

## RAPPORT DE RECHERCHE INTERNATIONALE

(article 18 et règles 43 et 44 du PCT)

Référence du dossier du déposant ou du mandataire B2928A - AD	<b>POUR SUITE</b> voir la notification de transmission du rapport de recherche internationale (formulaire PCT/ISA/220) et, le cas échéant, le point 5 ci-après <b>A DONNER</b>	
Demande internationale n° PCT/FR 97/00167	Date du dépôt international (jour/mois/année) 28/01/1997	(Date de priorité (la plus ancienne) (jour/mois/année) 29/01/1996
Déposant  BIOCEM et al.		

Le présent rapport de recherche internationale, établi par l'administration chargée de la recherche internationale, est transmis au déposant conformément à l'article 18. Une copie en est transmise au Bureau international.

Ce rapport de recherche internationale comprend 4 feuilles.

☒ Il est aussi accompagné d'une copie de chaque document relatif à l'état de la technique qui y est cité.

1. ☐ Il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (voir le cadre I).

2. ☐ Il y a absence d'unité de l'invention (voir le cadre II).

3. ☒ La demande internationale contient la divulgation d'un listage de séquence de nucléotides ou d'acides aminés et la recherche internationale a été effectuée sur la base du listage de séquence

☐ déposé avec la demande internationale

☒ fourni par le déposant séparément de la demande internationale

☐ sans être accompagnée d'une déclaration selon laquelle il n'inclut pas d'éléments allant au-delà de la divulgation faite dans la demande internationale telle qu'elle a été déposée.

☐ transcrit par l'administration

4. En ce qui concerne le titre, ☒ le texte est approuvé tel qu'il a été remis par le déposant.

☐ Le texte a été établi par l'administration et a la teneur suivante:

5. En ce qui concerne l'abrégé,

☒ le texte est approuvé tel qu'il a été remis par le déposant

☐ le texte (reproduit dans le cadre III) a été établi par l'administration conformément à la règle 38.2b). Le déposant peut présenter des observations à l'administration dans un délai d'un mois à compter de la date d'expédition du présent rapport de recherche internationale.

6. La figure des dessins à publier avec l'abrégé est la suivante:

Figure n°            ☐ suggérée par le déposant.

☐ parce que le déposant n'a pas suggéré de figure.

☐ parce que cette figure caractérise mieux l'invention.

☒ Aucune des figures n'est à publier.



# RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR 97/00167

<b>A. CLASSEMENT DE L'OBJET DE LA DEMANDE</b> CIB 6    C12N15/29    C12N15/82    C07K14/425    A01H5/00		
Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB		
<b>B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE</b> Documentation minimale consultée (système de classification suivi des symboles de classement) CIB 6    C12N    C07K    A01H		
Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche		
Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)		
<b>C. DOCUMENTS CONSIDERES COMME PERTINENTS</b>		
Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	WO 93 15221 A (HITACHI CHEMICAL CO LTD ;HITACHI CHEMICAL RESEARCH CENT (US)) 5 Août 1993 voir seq. I.D no. 186 voir page 46 ---	1
X	WO 93 12230 A (STANFORD RES INST INT) 24 Juin 1993 voir page 12, ligne 22; figure 14A voir page 65 ---	2
X	EMBL SEQUENCE DATABASE, REL.18, ACCESSION NO. M19784, 20-FEB-1989, XP002035349 FORNEY, J.D., ET AL.: voir séquence ---	2
-/--		
<div style="display: flex; justify-content: space-between;"> <span><input checked="" type="checkbox"/> Voir la suite du cadre C pour la fin de la liste des documents</span> <span><input checked="" type="checkbox"/> Les documents de familles de brevets sont indiqués en annexe</span> </div>		
<div style="display: flex;"> <div style="flex: 1;"> <p>* Catégories spéciales de documents cités:</p> <p>"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent</p> <p>"E" document antérieur, mais publié à la date de dépôt international ou après cette date</p> <p>"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)</p> <p>"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens</p> <p>"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée</p> </div> <div style="flex: 1;"> <p>"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention</p> <p>"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément</p> <p>"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier</p> <p>"&amp;" document qui fait partie de la même famille de brevets</p> </div> </div>		
Date à laquelle la recherche internationale a été effectivement achevée  <div style="text-align: center; font-size: 1.2em;">18 Juillet 1997</div>		Date d'expédition du présent rapport de recherche internationale  <div style="text-align: center; font-size: 1.2em;">3 0. 07. 97</div>
Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016		Fonctionnaire autorisé  <div style="text-align: center; font-size: 1.2em;">Maddox, A</div>





## C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	CHEMICAL ABSTRACTS, vol. 104, no. 19, 1986 Columbus, Ohio, US; abstract no. 168751, SCHOTT, H., ET AL.: "Synthesis of fragments of the terminal inverted repeating units of macronuclear DNA from hypotrichous ciliates" XP002034741 voir abrégé & MAKROMOL. CHEM., vol. 187, no. 1, 1986, pages 81-104, ---	3
X	EMBL SEQUENCE DATABASE ACCESSION NO. R40179, RELEASE 43, 8-05-1995., XP002034738 HILLIER, L., ET AL.: "yf70e09.s1 Homo sapiens cDNA clone 27461 3' similar to contains MER22 repetitive element" voir séquence ---	4
X	GCG-GENESEQ DATABASE ACCESSION NO. R28869, 23-MAR-1993, XP002034739 voir séquence & WO 92 19248 A (UNIVERSITY OF WASHINGTON) 12 Novembre 1992 ---	24
X	THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 270, no. 18, 5 Mai 1995, pages 10847-10854, XP002035350 QI, Z., ET AL.: "Reconstitution of neuronal cdc2-like kinase from bacteria-expressed cdk5 and an active fragment of the brain-specific activator" voir page 10852; tableau II ---	24
A	PLANT CELL 6 (12). 1994. 1911-1922., XP002017847 GELI M I ET AL: "Two structural domains mediate two sequential events in gamma-Zein targeting: Protein endoplasmic reticulum retention and protein body formation." cité dans la demande voir le document en entier ---	1-41
A	PLANT MOL BIOL 16 (1). 1991. 117-128., XP002017848 OHTANI T ET AL: "NORMAL AND LYSINE-CONTAINING ZEINS ARE UNSTABLE IN TRANSGENIC TOBACCO SEEDS." cité dans la demande voir le document en entier ---	1-41
	---	
	-/--	



## C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	EP 0 318 341 A (PLANT GENETIC SYSTEMS NV) 31 Mai 1989 voir le document en entier ---	1-41
A	EP 0 208 418 A (LUBRIZOL GENETICS INC) 14 Janvier 1987 voir le document en entier ---	1-41
A	WO 95 06128 A (DEKALB GENETICS CORP) 2 Mars 1995 voir page 211 - page 222 ---	41
T	PLANT MOLECULAR BIOLOGY, vol. 34, Mai 1997, pages 139-149, XP002034740 TORRENT, M., ET AL.: "Lysine-rich modified gamma-zeins accumulate in protein bodies of transiently transformed maize endosperms" voir le document en entier -----	1-41



## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 97/00167

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9315221 A	05-08-93	JP 7506482 T	20-07-95
WO 9312230 A	24-06-93	WO 9312234 A	24-06-93
EP 0318341 A	31-05-89	CA 1337048 A	19-09-95
		AU 634987 B	11-03-93
		AU 4495189 A	01-05-90
		CA 2000661 A	14-04-90
		WO 9004032 A	19-04-90
		JP 3502644 T	20-06-91
		US 5589615 A	31-12-96
		AU 2811889 A	23-05-89
		WO 8903887 A	05-05-89
		EP 0319353 A	07-06-89
		JP 2501802 T	21-06-90
		US 5487991 A	30-01-96
		AT 140959 T	15-08-96
		DE 3855455 D	05-09-96
		DE 3855455 T	09-01-97
		AT 143694 T	15-10-96
		DE 3855591 D	07-11-96
		DE 3855591 T	03-04-97
		EP 0723019 A	24-07-96
		US 5623067 A	22-04-97
EP 0208418 A	14-01-87	AU 592090 B	04-01-90
		AU 5853686 A	18-12-86
		CA 1304025 A	23-06-92
		DE 3682233 A	05-12-91
		JP 62059300 A	14-03-87
		US 4885357 A	05-12-89
		US 4886878 A	12-12-89
WO 9506128 A	02-03-95	AU 7716994 A	21-03-95
		CA 2170260 A	02-03-95
		EP 0721509 A	17-07-96
		HU 74392 A	30-12-96
		ZA 9406488 A	30-11-95



## INTERNATIONAL SEARCH REPORT

 International application No.  
 PCT/US93/00977

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Gene Analysis Techniques, Volume 4, issued 1987, C. B. Harley, "Hybridization of Oligo(dT) to RNA on Nitrocellulose", pages 17-22, see entire document.	1-28, 33-40
X Y	Methods in Enzymology, Vol. 65, Pt. 1, issued 1980, A. R. Dunn et al, "Mapping Viral mRNAs by Sandwich Hybridization", pages 468-478, especially page 46, second full paragraph.	<u>1-3, 6</u> 4, 5, 7-28, 33-40
X Y	DNA, Volume 7, No. 2, issued 1988, A. Palva et al, "Laboratory Methods: Quantification of $\alpha$ -Amylase mRNA in <u>Bacillus subtilis</u> by Nucleic Acid Sandwich Hybridization", pages 135-142, see entire document.	<u>1, 2, 6</u> 4, 5, 7-28, 33-40
X Y	US, A, 4,957,865 (Samarut et al) 18 September 1990, col. 11, lines 34-41.	<u>29-31</u> 32
Y	Journal of Molecular Evolution, Volume 19, Numbers 3-4, R. Nussinov, "Efficient algorithms for searching for exact repetition of nucleotide sequences.", pages 283-285, see Abstract 05061573.	23
Y	Computer Applications in the Biosciences, Volume 8, No. 3, issued September 1987, W. W. Ralph et al, "A modified Chou and Fasman protein structure algorithm", pages 211-216, see Abstract 06625157.	23-25
Y	Biochemical and Biophysical Research Communications, Volume 170, No. 1, issued 1990, C. Hough et al, "Differential down-regulation of $\beta_1$ - and $\beta_2$ -adrenergic receptor mRNA in C <sub>6</sub> glioma cells", pages 46-52, see entire document.	9, 33, 34
Y	Analytical Biochemistry, Volume 196, issued 1991, A. Ballagi-Pordany et al, "Quantitative determination of mRNA phenotypes by the polymerase chain reaction", pages 89-94, see especially the Abstract.	1-3
Y	Journal of Biological Chemistry, Volume 265, No. 27, issued 25 September 1990, B. Feve et al, "Differential regulation of $\beta_1$ - and $\beta_2$ -adrenergic receptor protein and mRNA levels by glucocorticoids during 3T3-F442A adipose differentiation", pages 16343-16349, see entire document.	9, 33, 34
Y	Journal of Biological Chemistry, Volume 266, No. 35, issued 15 December 1991, P. Muzzin et al, "An adipose tissue-specific $\beta$ -adrenergic receptor", pages 24053-24058, especially Figures 1 and 2.	9, 33, 34

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Analytical Biochemistry, Volume 170, issued 1988, C. Albretsen et al, "Optimal conditions for hybridization with oligonucleotides: A study with <u>myc</u> -oncogene DNA probes", pages 193-202, especially paragraph bridging pages 193 and 194.	1-28
Y	Proceedings of the National Academy of Science USA, Volume 84, issued March 1987, M. S. Waterman et al, "Phase transitions in sequence matches and nucleic acid structure", pages 1239-1243, see entire document.	1-28
Y	Journal of Molecular Biology, Volume 197, issued 1987, M. S. Waterman et al, "A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons", pages 723-728, see entire document.	1-23
Y	Biotechniques, Volume 12, No. 2, issued Feb 1992, R. J. Cano et al, "DNA hybridization assay using ATTOPHOS <sup>TM</sup> , a fluorescent substrate for alkaline phosphatase", pages 264-269, see entire document.	1, 2, 18, 21, 22
Y	Genomics, Vol. 11, issued 1991, W. R. Pearson, "Searching protein sequence libraries: Comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms", pages 635-650, see entire document.	1-23
Y	Biochemistry, Vol. 30, issued 1991, N. P. Gerard, "Human substance P receptor (NK-1): Organization of the gene, chromosome localization, and functional expression of cDNA clones", pages 10640-10646, especially Figure 1.	12, 39, 40
Y	The EMBO Journal, Volume 8, No. 5, issued 1989, S.-I. Hirai et al, "Characterization of <u>junD</u> : a new member of the <u>jun</u> proto-oncogene family", pages 1433-1439, especially Figure 1b.	35, 36
Y	Journal of Biological Chemistry, Volume 262, No. 29, issued 15 October 1987, D. T. Jones et al, "Molecular cloning of five GTP-binding protein cDNA species from rat olfactory neuroepithelium", pages 14241-14249, especially Figures 2-6.	10, 11, 13-15, 37-38
Y	Journal of Biological Chemistry, Volume 265, No. 3, issued 25 January 1990, L. Li et al, "Different members of the <u>jun</u> proto-oncogene family exhibit distinct patterns of expression in response to type $\beta$ transforming growth factor", pages 1556-1562, especially Figure 5.	35, 36



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Journal of Biological Chemistry, Volume 264, No. 30, issued 25 October 1989, Y. Yokota et al, "Molecular characterization of a functional cDNA for rat substance P receptor", pages 17649-17652, especially Figure 2.	12, 39, 40
Y	Journal of Biological Chemistry, Volume 263, No. 14, issued 15 May 1988, J. Codina et al, " $\alpha_i$ -3 cDNA encodes the $\alpha$ subunit of $G_k$ , the stimulatory G protein of receptor-regulated $K^+$ channels", pages 6746-6750, especially Figure 2.	13, 37, 38
Y	Journal of Biological Chemistry, Volume 263, No. 14, issued 15 May 1988, H. Itoh et al, "Presence of three distinct molecular species of $G_i$ protein $\alpha$ subunit", pages 6656-6664, especially Figures 2, 4-6, and 8.	10, 11, 13-15, 37, 38
Y	Proceedings of the National Academy of Science USA, Volume 85, issued December 1988, K. Hattori et al, "Structure and chromosomal localization of the functional intronless human <u>JUN</u> protooncogene", pages 9148-9152, especially Figure 5.	35, 36
Y	Proceedings of the National Academy of Science USA, Volume 84, issued August 1987, P. Bray et al, "Human cDNA clones for an $\alpha$ subunit of $G_i$ signal-transduction protein", pages 5115-5119, especially Figure 2 and 5-7.	10, 11, 13-15, 37, 38
Y	Proceedings of the National Academy of Science USA, Volume 84, issued November 1987, C. R. Beals et al, "A small multigene family encodes $G_i$ signal-transduction proteins", pages 7886-7890, especially Figures 1, 2, and 4.	10, 11, 13-15, 37, 38
Y	Nucleic Acids Research, Volume 16, No. 8, issued 1988, B. A. Harris, "Complete cDNA sequence of a human stimulatory GTP-binding protein alpha subunit", page 3585.	10, 11, 13-15, 37, 38
Y	FEBS Letters, Volume 206, No. 1, issued September 1986, R. Mattera et al, "Identification by molecular cloning of two forms of the $\alpha$ -subunit of the human liver stimulatory ( $G_i$ ) regulatory component of adenylyl cyclase", pages 36-42, especially Figure 2.	10, 11, 13-15, 37, 38
Y	FEBS Letters, Volume 219, No. 1, issued July 1987, J. R. Didsbury et al, "Molecular cloning of a new human G protein", pages 259-263, especially Figures 1 and 2.	10, 11, 13-15, 37, 38
Y	Nucleic Acids Research, Volume 19, No. 17, issued 1991, A. Swaroop et al, "Differential expression of novel $G_{s\alpha}$ signal transduction protein cDNA species", pages 4725-4729, see entire document.	10, 11, 13-15, 37, 38

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

N-GENESEQ, GENBANK, GENBANK-NEW, UEMBL, EMBL-NEW, CA, MEDLINE, BIOSIS, APS, search terms: Waterman, Pevzner, second hybridization strength, continuous seed filtration, mitsuhashi probe selection, calculation of melting temperature of nucleic acids, H-site model, nucleation threshold, mismatch model, measuring or assaying beta receptor mRNA, measuring or assaying substance P mRNA, measuring or assaying jun mRNA, measuring or assaying Gs, Go, or Gi protein mRNA

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-9, 16-28, 33 and 34, drawn to a method for detecting and quantifying specific mRNAs and a polynucleotide immobilized support for use in said method which detects  $\beta$ -receptor mRNA, classified in Class 435/6, for example.
- II. Claim 10, drawn to a method for detecting and quantifying specific Gs protein mRNA, classified in Class 435/6, for example.
- III. Claim 11, drawn to a method for detecting and quantifying specific Gi-2 protein mRNA, classified in Class 435/6, for example.
- IV. Claims 12, 39, and 40, drawn to a method for detecting and quantifying specific substance P mRNA and polynucleotide support for accomplishing the method, classified in Class 435/6, for example.
- V. Claim 13, drawn to a method for detecting and quantifying specific Gi-3 protein mRNA, classified in Class 435/6, for example.
- VI. Claim 14, drawn to a method for detecting and quantifying specific Gi-1 protein mRNA, classified in Class 435/6, for example.
- VII. Claims 15, 37, and 38, drawn to a method for detecting and quantifying specific G protein mRNA and polynucleotide immobilized support for accomplishing the method, classified in Class 435/6 for example.
- VIII. Claims 35 and 36, drawn to a polynucleotide immobilized support for the detection and quantification of jun-oncogene mRNA, classified in Class 536/24.3.
- IX. Claims 29-32, drawn to a method of inhibiting RNase activity in a sample, classified in Class 435/19.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

(Telephone practice)  
Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark a Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

(Telephone practice)  
Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>5</sup> : <b>C12Q 1/44, 1/68, C07H 21/00</b>		<b>A1</b>	(11) International Publication Number: <b>WO 93/15221</b>
			(43) International Publication Date: 5 August 1993 (05.08.93)

<p>(21) International Application Number: PCT/US93/00977</p> <p>(22) International Filing Date: 29 January 1993 (29.01.93)</p> <p>(30) Priority data:</p> <table border="0"> <tr> <td>07/827,208</td> <td>29 January 1992 (29.01.92)</td> <td>US</td> </tr> <tr> <td>07/857,059</td> <td>24 March 1992 (24.03.92)</td> <td>US</td> </tr> <tr> <td>07/974,409</td> <td>12 November 1992 (12.11.92)</td> <td>US</td> </tr> </table> <p>(60) Parent Applications or Grants</p> <p>(63) Related by Continuation</p> <table border="0"> <tr> <td>US</td> <td>07/974,409 (CIP)</td> </tr> <tr> <td>Filed on</td> <td>12 November 1992 (12.11.92)</td> </tr> <tr> <td>US</td> <td>07/857,059 (CIP)</td> </tr> <tr> <td>Filed on</td> <td>24 March 1992 (24.03.92)</td> </tr> <tr> <td>US</td> <td>07/827,208 (CIP)</td> </tr> <tr> <td>Filed on</td> <td>29 January 1992 (29.01.92)</td> </tr> </table> <p>(71) Applicants (for all designated States except US): HITACHI CHEMICAL CO., LTD. [JP/JP]; 48 Wadai, Tsukuba-City, Ibaraki (JP). HITACHI CHEMICAL RESEARCH CENTER, INC. [US/US]; 1003 Health Sciences Road West, Irvine, CA 92715 (US).</p>	07/827,208	29 January 1992 (29.01.92)	US	07/857,059	24 March 1992 (24.03.92)	US	07/974,409	12 November 1992 (12.11.92)	US	US	07/974,409 (CIP)	Filed on	12 November 1992 (12.11.92)	US	07/857,059 (CIP)	Filed on	24 March 1992 (24.03.92)	US	07/827,208 (CIP)	Filed on	29 January 1992 (29.01.92)	<p>(72) Inventors; and (75) Inventors/Applicants (for US only): AKITAYA, Tatsuo [JP/JP]; 20-2, Hirasawa Terasu, 16-15-2, 1-chome, Takasuzu-machi, Hitachi Ibaraki (JP). COOPER, Allan [US/US]; 3607 N.E. 172nd Avenue, Bellview, WA (US). MITSUHASHI, Masato [JP/US]; 8 Brookmont, Irvine, CA 92714 (US).</p> <p>(74) Agents: ALTMAN, Daniel, E. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, 16th Floor, Newport Beach, CA 92660 (US).</p> <p>(81) Designated States: CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p><b>Published</b> <i>With international search report.</i></p>
07/827,208	29 January 1992 (29.01.92)	US																				
07/857,059	24 March 1992 (24.03.92)	US																				
07/974,409	12 November 1992 (12.11.92)	US																				
US	07/974,409 (CIP)																					
Filed on	12 November 1992 (12.11.92)																					
US	07/857,059 (CIP)																					
Filed on	24 March 1992 (24.03.92)																					
US	07/827,208 (CIP)																					
Filed on	29 January 1992 (29.01.92)																					

(54) Title: METHOD FOR MEASURING MESSENGER RNA

(57) Abstract

The present invention provides a method for detecting and quantifying mRNA in a sample. The mRNA that can be detected has a unique sequence. The method includes immobilizing a first polynucleotide to an insoluble support. The first polynucleotide has a first sequence that hybridizes to the unique sequence on the mRNA. After immobilization of the first polynucleotide, the sample is applied to the insoluble support under conditions that allow the unique sequence on the mRNA to hybridize with the first polynucleotide. Thereafter, a labeled probe can be applied to the insoluble support. This labeled probe has a sequence thereon that hybridizes to a portion of the mRNA other than the unique sequence. The application of the labeled probe is performed under conditions that allow the labeled probe to hybridize with mRNA immobilized on said support, if present. Finally, the amount of the labeled probe immobilized on the support is measured to provide an indication of the amount of mRNA present in the sample.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				



**METHOD FOR MEASURING MESSENGER RNA**Background of the InventionField of the Invention

5 This invention is related to a method for measuring messenger RNA (mRNA) and polynucleotide-immobilized supports useful in this method. Specifically, this invention is related to measuring methods and reagents in which various types of mRNA can be measured at the same time. In one embodiment of this invention, the mRNA need not be purified from the cell, and the mRNA can be rapidly quantified with high sensitivity.

Description of Previous Techniques

10 Previously well-known techniques for measuring mRNA levels include the Northern Blot Method (See, e.g., J. Sambrook, et al., Molecular Cloning, 2nd ed., Cold spring Harbor Laboratory Press, 1989 hereafter referred to as "Molecular Cloning"), the Dot Blot Method (See, e.g., Molecular Cloning), Ribonuclease Protection Assay (See, e.g., Molecular Cloning), and the Reverse PCR Method (H.A. Erlich (ed.), PCR Technology-Principles and  
15 Applications for DNA Amplification, Stockton Press, New York, NY, 1989). The disclosures of all of these previous references are hereby incorporated by reference.

The Northern Blot Method detects mRNA which is separated by molecular weight using electrophoresis. The separated molecules are immobilized on a membrane and mRNA levels are determined by hybridization of marker DNA probes. With this method,  
20 the molecular weight of the subject mRNA must generally be known. By using this method a complex between a labeled probe, mRNA and an insoluble membrane is formed. The amount of label in the complex at the appropriate site on the membrane is then measured colorimetrically or by radioactive emissions.

The Dot Blot Method detects mRNA by hybridization with a labeled probe after  
25 purified mRNA is immobilized as a spot on a membrane. This method can be used to rapidly examine many samples at one time. However, both the Dot Blot and the Northern Blot methods require the use of a specific labeled probe for each mRNA which is desired to be detected.

The Ribonuclease Protection Assay method is used to detect mRNA after it is  
30 initially purified. The purified mRNA is hybridized with a labeled RNA probe, and then treated with a ribonuclease that specifically digests single stranded RNA. The labeled RNA probe will form a duplex in the solution and be protected from digestion. This is then immobilized on a membrane after separation of the double-stranded RNA/probe complex.

The reverse PCR method is used where cDNA is synthesized using reverse transcriptase and purified mRNA. This cDNA can be immobilized to a membrane. The synthesized cDNA is then amplified using standard PCR (Polymerase Chain Reaction) methods. After separating cDNA according to its molecular weight by electrophoresis, it becomes a single band. The DNA is then immobilized on a membrane, and detected by hybridization with a labeled DNA probe. However, after PCR, no quantitative information remains due to the amplification of the original message.

All of the foregoing methods take extended periods of time to perform since the mRNA used must be in at least a partially purified form. Also, electrophoresis of the samples and immobilization to a membrane is necessary. Both of these steps reduce the reliability of quantification of the mRNA.

Thus, there remains a need to resolve the above problems, and to provide a method of measuring for rapidly determining messenger RNA levels and a measuring reagent useful in such a method.

15

#### Summary of the Invention

In the mRNA assay methods of the present invention, a first and a second polynucleotide probe can be used. The inventors have discovered that immobilization of probes (a first polynucleotide) rather than samples, on a water-insoluble support can provide a basis for the rapid quantification of multiple unpurified samples containing mRNA. Applicants have provided an efficient method of identifying appropriate specific probes for this purpose. In preferred forms of the present invention, the first polynucleotide probe has a nucleotide sequence complementary to a region which is unique in the sample to the mRNA and does not bind other portions in the target mRNA. The inventors also discovered that a single kind of second polynucleotide probe can be used for simultaneous quantification of various mRNAs. However, any of a variety of probes that bind to a different region of the mRNA than the first probe can be used for this purpose. In this regard, the inventors have found that a second polynucleotide probe containing sequences complementary to a polyadenylic acid tail of mRNA can be used in such a simultaneous assay method. Based on these and other discoveries, the inventors established the present invention.

One embodiment of the present invention is a highly sensitive, quantitative and rapid method for detecting and quantifying mRNA in a sample without the need to purify mRNA from cells. In this embodiment of the invention, the method comprises the following steps:

- (a) identifying a polynucleotide sequence that is unique to the mRNA;

(b) immobilizing a first polynucleotide to an insoluble support, the first polynucleotide having a first sequence complementary to the sequence unique to the mRNA;

(c) incubating the sample with the insoluble support under conditions wherein the unique sequence will hybridize with the first polynucleotide, thereby immobilizing mRNA present in the sample to the insoluble support;

(d) washing non-immobilized components of the sample from the insoluble support;

(e) labeling mRNA on the support in a manner that label is incorporated onto the support related to the amount of mRNA on the support; and

(f) measuring the amount of label immobilized on the support.

In a preferred embodiment, the above step of labeling mRNA on the support involves incubating a labeled probe, most preferably poly-d(T), with the insoluble support under conditions wherein the labeled probe will hybridize with mRNA on the support, and the labeled probe bears a label and is complementary to a part of the mRNA other than a part of the mRNA complementary to the first probe, thereby immobilizing to the insoluble support the label on the labeled probe which has hybridized with the mRNA; and thereafter washing any non-immobilized labeled probe from the insoluble support.

In addition, the above method of labeling nucleic acids in step (e) can preferably involve labeling the oligonucleotides with a nucleic acid stain, most preferably from the group consisting of ethidium bromide, yoyo-1 and toto-1.

In a further preferred embodiment, the above method can utilize a label, wherein the label comprises a radionuclide, or alternatively biotin. In the method using biotin, the incubation can preferably be with enzyme-conjugated streptavidin; followed by measurement of the amount of enzyme bound to the insoluble support. In this embodiment step (f) can additionally comprise adding streptavidin bound to alkaline phosphatase to the insoluble support.

In still another embodiment of the preferred method, the mRNA can code for a  $\beta$  receptor, and wherein the first polynucleotide comprises the sequence of SEQ ID NO:6. An alternative embodiment of the above method includes mRNA coding for the  $\alpha$  subunit of Gs protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:1, 145-146, 209-212, 268-287, 705 and 708.

Yet another alternative mRNA sequence that can be used with the present invention method codes for the Gi-2 protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:703, 806, 142, 143, 187-197 and 238-253.

5 The present method can also preferably involve the use mRNA coding for substance P receptor, wherein the first polynucleotide comprises the sequence of SEQ ID NO:5.

mRNA encoding Gi-3 protein can also be used in the present method, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:704, 707, 144, 198-208 and 254-267.

10 Still another embodiment of the above method can involve the mRNA coding for the Gi-1 protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:702, 805, 141, and 224-237. An even further preferred embodiment includes mRNA coding for G protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:700, 701, 139, 140, 148-159, 162-172, 176-186 and 3-4.

15 Most preferably, the first sequence hybridizes to a unique sequence of the mRNA and not to other polynucleotides present in the sample, and even more preferably, the sample comprises a cell lysate.

20 In an alternative preferred embodiment, the label in the above method can be measured by light emitted therefrom, and step (f) of the method comprises measuring the amount of light emitted by the label. The method of measuring the light can preferably involve; recording the amount of light on film; and measuring the exposure of this film using a densitometer. In addition, the light measurement can alternatively comprise adding ATTOPHOS and measuring fluorescence emitted using a fluorimeter.

25 The present invention method in step (a) above can advantageously include a computer program, with the computer program preferably involving use of an H-site model. The H-site model of step (a) is preferably used to identify the first polynucleotide probe of the method by the steps of:

30 specifying a minimum melting temperature for the first nucleotide probe and the nucleotide sequence specific to the organism;  
specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures ( $T_m$ ) of the first nucleotide probe and the sequence specific to the organism at every possible hybridization point; and

selecting the nucleotide probe having the highest  $T_m$  value.

5 The computerized method of identifying the first polynucleotide probe can preferably have the melting temperature determined by the formula;

$T_m = 81.5 - 16.6(\log[\text{Na}]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$ , wherein  $\log[\text{Na}]$  is the log function of the sodium concentration,  $0.063\%(\text{formamide})$  is the concentration of formamide,  $\%(G+C)$  is the percentage of matched GC base pairs, and  $N$  is the probe  
10 length.

Additionally the method of step (a) can be used to identify the labeled probe referred to above using the H-site computerized model by the steps of:

specifying a minimum melting temperature for the labeled probe and the nucleotide sequence specific to the organism;

15 specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures ( $T_m$ ) of the first nucleotide probe and the sequence specific to the organism at every possible hybridization point; and

20 selecting the nucleotide probe of the proper length having the lowest  $T_m$  value.

A method of identifying the second nucleotide probe can involve determining the melting temperature by the formula:

$T_m = 81.5 - 16.6(\log[\text{Na}]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$ , wherein  $\log[\text{Na}]$  is the log function of the sodium concentration,  $0.063\%(\text{formamide})$  is the concentration of formamide,  $\%(G+C)$  is the percentage of matched GC base pairs, and  $N$  is the probe  
25 length.

A different embodiment of the present invention is a method of inhibiting RNase activity in a sample containing RNA in a buffer system containing SDS and EDTA, comprising adding vanadyl ribonucleoside complex to the sample.  
30

Preferably, there is at least a five-fold molar excess of EDTA over RNA in the sample and most preferably, the vanadyl ribonucleoside complex is added in combination with proteinase K. In an alternative preferred embodiment, the vanadyl ribonucleoside complex is added in combination with RNasin.

Yet another different embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of  $\beta$  receptor mRNA in a sample, comprising:

- an insoluble support;
- 5 a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to  $\beta$  receptor mRNA.

More preferably, the sequence unique to  $\beta$  receptor mRNA comprises a sequence selected from the group consisting of SEQ ID NO:6.

- 10 Still another preferred embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of jun oncogene mRNA in a sample, comprising:

- an insoluble support;
- 15 a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to jun oncogene mRNA.

Preferably, the sequence unique to the jun oncogene mRNA comprises a sequence selected from the group consisting of SEQ ID NO:11-35, 39-53, 62-138, 310-342, 600-606, 614-620, and 628-634.

- 20 Another embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of G protein mRNA in a sample, comprising:

- an insoluble support;
- 25 a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to G protein mRNA.

Preferably, the sequence unique to G protein mRNA comprises a sequence selected from the group consisting of 1, 2, 700-709, 805, 806, 139-159, 162-172 and 176-309.

- 30 Another preferred embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of Substance P receptor mRNA in a sample, comprising:

- an insoluble support;

a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to Substance P receptor mRNA.

Preferably the sequence unique to Substance P receptor mRNA comprises the sequence of  
5 SEQ ID NO:5.

#### Brief Description of the Figures

Figure 1 illustrates an immobilization procedure of a first polynucleotide probe onto an insoluble support by the maleimide method.

10 Figure 2 illustrates an immobilization procedure of a first nucleotide probe onto an insoluble support by the carbodiimide method.

Figure 3 schematically shows the principle of an mRNA quantification method in accordance with the present invention.

15 Figure 4A shows a photographic image of an X-ray film exposed to chemiluminescent light derived from the human  $\beta_2$  adrenergic receptor cDNA probe on a Northern membrane, in which various concentrations of the human  $\beta_2$  adrenergic receptor mRNA were immobilized. 4B shows a photographic image of an X-ray film exposed to chemiluminescent light derived from the oligo-(dT) probe in accordance with the present invention in which various concentrations of the human  $\beta_2$  adrenergic receptor mRNA were trapped onto the first oligonucleotide-immobilized plastic plate. 4C shows a densitogram  
20 of the image shown in 4B. 4D graphically depicts the relationship between relative intensity of chemiluminescent signals and concentrations of mRNA of the image shown in 4B.

Figure 5 graphically depicts the linear relationship between relative intensity of chemiluminescent signals and concentrations of mRNA in a separate experiment from that shown in Figures 4B-D.

25 Figure 6 graphically depicts a calibration line for the rat Gs protein mRNA of the invented method.

Figure 7 shows a photographic image of a polaroid film exposed to chemiluminescent light derived from oligo-(dT) probe by the invented method, in which Gi-2 and GS protein mRNA in various cells were trapped onto the first polynucleotide-immobilized plastic plate.

30 Figure 8 shows a photographic image of an X-ray film exposed to chemiluminescent light derived from the Gi-2 and Gs protein cDNA probes on the Southern membrane, in which PCR-amplified G protein specific DNA from 4 different cell lines was immobilized.

Figure 9 graphically depicts a calibration line for the human substance P receptor mRNA of the invented method.

Figure 10 shows a gel showing the effect of various RNase inhibitors on mRNA preparations containing SDS and EDTA.

Figure 11 shows a gel of five different G protein oligonucleotides amplified with the G<sub>2</sub> and G<sub>4</sub> PCR primers and also provides Southern blots using each of the five G protein sequences as a probe.

Figure 12 shows a gel of samples containing varying amounts of each of five different G protein oligonucleotides (as indicated by number of "+" symbols) amplified with the G<sub>2</sub> and G<sub>4</sub> PCR primers and also provides Southern blots using each of the five G protein sequences as a probe.

Figure 13 shows a gel of  $\lambda$ ZAP cDNA libraries from rat pituitary (P), kidney (K) and intestinal (I) amplified with each of five different G protein primers, as indicated.

Figure 14 shows a gel of 500 bp DNA from cDNAs of human IM9 and Jurkat cells amplified with G<sub>2</sub> and G<sub>4</sub> PCR primers.

Figure 15 is a graph showing the relationship between YOYO-1 fluorescence and amount of immobilized oligonucleotide.

Figure 16 is a graph showing the time course of YOYO-1 analysis.

Figure 17 is a graph showing the dose response of YOYO-1 concentration.

Figure 18 is a bar graph showing the reproducibility of YOYO-1 staining based on quantitation of immobilized oligonucleotides.

Figure 19 exhibits a Polaroid picture of agarose gel electrophoresis of the PCR amplified ds-cDNA of jun oncogene from human peripheral blood leukocytes.

Figure 20 is a simplified block diagram of a computer system illustrating the overall design of this invention;

Figures 21A-21C show display screen representations of the main oligoprobe design station dialog windows of this invention;

Figures 22A and 22B are flow charts of the overall invention illustrating the program and the invention's sequence and structure;

Figure 23 is a display screen representation of the Mitsuhashi probe selection diagram;

Figure 24 is a display screen representation of the probeinfo and matchinfo window;

Figure 25 is a display screen representation of the probesedit window;

Figures 26A and 26B are printouts of the probesedit output file;

Figure 27 is a flow chart of the overall k<sub>diff</sub> program of the Mismatch Model of this invention, including its sequence and structure:



Figures 28A and 28B are flow charts of the k\_diff module of this invention;

Figure 29 is a flow chart of the hashing module of this invention;

Figure 30 is a flow chart of the tran module of this invention;

Figure 31 is a flow chart of the let\_dig module of this invention;

5 Figure 32 is a flow chart of the update module of this invention;

Figures 33A and 33B are flow charts of the assembly module of this invention;

Figures 34A and 34B are flow charts of the seqload module of this invention;

Figures 35A and 35B are flow charts of the read1 module of this invention;

Figure 36 is a flow chart of the dig\_let module of this invention;

10 Figures 37A and 37B are flow charts of the q\_colour module of this invention;

Figure 38 is a flow chart of the hit\_ext module of this invention;

Figure 39 is a flow chart of the colour module of this invention;

Figure 40 is the first page of a printout of a sample file containing the output of the Mismatch Model program of this invention;

15 Figure 41 is a flow chart of the H-Site Model, stage I, covering the creation of a preprocessed preparation file of this invention;

Figure 42 is a flow chart of the H-Site Model, stage II, covering the preparation of the target sequence(s);

20 Figures 43A-43D are flow charts of the H-Site Model, stage III, covering the calculation of MPSD data;

Figure 44A is the first page of a printout of a sample file containing output of Mismatch Model program;

Figure 44B is the first page of a printout of a sample file containing output of H-Site Model program;

25 Figure 45 is a flow chart of the processing used to create the Mitsuhashi probe selection diagram (MPSD);

Figure 46 is a flow chart of processing used to create the matchinfo window;

Figure 47 is a printout of a sample target species file;

Figures 48A-48C are printouts of a sample preparation file.

30 Detailed Description of the Invention

#### Target mRNA

Previously, in order to genetically diagnose an infectious disease, procedures utilizing DNA were generally performed. Although, with DNA diagnosis, the existence of an external

gene can be confirmed, it cannot be determined if the known gene is in an unexpressed carrier state or expressed in the diseased state.

In living organisms, once a gene is activated, the DNA coding for that gene is transcribed into mRNA, and the mRNA is subsequently translated into protein. Thus, the activation of the external gene can be determined by measuring the levels of mRNA.

Currently, to genetically diagnose a protein deficiency state, most experimenters have limited themselves to analysis at the DNA level. However, this method cannot be used to determine abnormalities of transcription or translation. If mRNA levels could be measured easily, the diseased state might be more precisely recognized. In a diseased state where mRNA cannot be detected, it can be said that there is a defect in transcription, while if normal levels of mRNA are detected, there is probably an abnormality in translation.

In addition, the amount of mRNA produced in the cell reflects the amount of expression of a gene, so it would be useful in the course of an abnormal condition, such as an illness or drug reaction, to examine mRNA levels. As discussed in greater detail below, measuring any of a large number of mRNAs can provide clinically useful information. Particular examples include mRNA for the  $\beta$  receptor, substance P, G protein and oncogenes, such as jun.

For the aforementioned reasons, the ability to measure mRNA would be useful for diagnosing and recognizing the pathophysiology of various disease states. This method would be especially useful for diagnosis of hereditary diseases, cancer, and infectious diseases.

A preferred target of this invention is mRNA of eukaryote cells which contain polyadenylic acid tail on their 3'-terminal end. Once the specific nucleotide sequence of the mRNA is known, virtually any mRNA can be measured using the methods of the present invention. Examples of such mRNAs are human G protein, human  $\beta_2$  receptor, substance P, lymphocyte surface antigen, immunoglobulin, collagen, and adrenaline receptor.

#### **Test Samples**

Test samples containing target mRNA can be obtained from a variety of sources. Preferred sources include living samples or cell solutions which have been treated to inactivate ribonuclease (RNase) activity. Purification of mRNA for standard samples can be conducted according to the methods usually associated with Northern Blot procedures or Dot Blot procedures method. Kits for the purification of mRNA are commercially available.

A test sample can be prepared from a living sample by treatment with 10 mM of ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 0.2 M NaCl, 0.5% of sodium dodecyl sulfate (SDS), 500 Unit/ml of RNase inhibitor, 10mM of Vanadyl Ribonucleoside Complex and 200 µg/l of Proteinase K (hereafter called Lysis buffer) to lyse the cells and inhibit  
5 RNase activity. After lysis of the cells, NaCl concentration is adjusted to 0.5M.

In the methods of the present invention that involve probing samples containing RNA, or in any procedure in which it is desired to prevent the degradation of RNA, it is advantageous to inhibit the activity of any ribonucleases (RNases) which may be present. One Rnase inhibitor which can be used is Vanadyl Ribonucleoside Complex (VRC)  
10 (Bethesda Research Laboratories, Gaithersburg, MD). VRC has been reported to be useful during cell fractionation and in the preparation of RNA, and has been shown not to interfere with the phenol extraction or ethanol precipitation of RNA. In addition, VRC does not affect other cytoplasmic components of cells. Therefore, VRC is an ideal inhibitor of RNase in many experimental procedures.

However, prior art procedures for inhibiting RNases with VRC taught that VRC  
15 should not be used in buffer systems containing EDTA or SDS, which are commonly used in the field of molecular biology. The reason for this prohibition was that it was believed that the Complex would dissociate in the presence of these buffers, leading to a loss of RNase inhibiting activity. In fact, the BRL insert accompanying the VRC product  
20 recommends that a five- to ten-fold molar excess of EDTA be added to an RNA solution containing VRC in order "destroy" the VRC prior to ethanol extraction of the RNA from the solution. The apparent inability to use VRC together with common buffers that include EDTA and SDS thus presented a major impediment to the exploitation of VRC as an RNase inhibitor.

We have discovered, however, that VRC is in fact an effective RNase inhibitor even  
25 in the presence of SDS and/or EDTA. Thus, VRC can be used in assays which use buffers including SDS or EDTA, where heretofore it was believed that VRC would not be effective in such systems. VRC is an effective RNase inhibitor at the concentrations of EDTA and SDS that are normally used when manipulating RNA or when performing a variety of other  
30 molecular biology techniques. For example, we have found that VRC effectively inhibits RNase in a buffer solution comprising approximately 1 mM EDTA. We also found VRC to be effective in .5% SDS solutions, and is believed to be effective in solutions ranging up to 2% SDS, more preferably up to 1% SDS. VRC can, of course, also be used in other solutions including EDTA and SDS.

We have found that VRC is a particularly potent inhibitor of RNases when used in combination with Proteinase K. Proteinase K is also available from BRL. As shown in the gel in Figure 10, mRNA prepared from U937 cells (human macrophage cell line) was protected from RNase degradation by a combination of VRC and Proteinase K. Lane 1 of the gel shows the mRNA from a cell preparation which included VRC, Proteinase K, and RNasin, while lane 2 represents the mRNA from a cell preparation that included VRC and Proteinase K. RNasin is available from Promega of Madison, WI. The distinct band 10 in lanes 1 and 2 matches the band seen in lane 7, which contains pure clonal cDNA from an RNase-free preparation of U937 cells, thus showing that the mRNA in the preparations of lanes 1 and 2 did not experience substantial mRNA degradation.

A comparison of lanes 1 and 2 with lanes 3-6 shows that VRC in combination with Proteinase K inhibits RNase activity in the above-mentioned mRNA preparation from U937 cells to a far greater extent than either Protein K alone (lane 4). Proteinase K in combination with RNasin (lane 3), or a commercial RNase-inhibiting preparation sold under the name "FastTrack" (available from In Vitrogen of San Diego, CA) (lane 5). None of lanes 3-5 exhibit the distinct band 10 representing undegraded mRNA. On the contrary, lane 4 (Proteinase K alone) and lane 5 (FastTrack) have the same smear of degraded mRNA as band 20 in lane 6 (no inhibitors). The gel shown in Figure 10 also shows the effectiveness of VRC in a buffer solution of 1 mM EDTA and .5% SDS, which is the buffer used in the U937 cell preparations tested, since lane 2 (VRC and Proteinase K) shows less mRNA degradation than lane 4 (Proteinase K alone) or lane 3 (Proteinase K and RNasin).

#### **First Polynucleotide Probe**

In the practice of the present invention, a first polynucleotide probe is bound to an insoluble support. This first polynucleotide is used to trap the target mRNA to a solid support. Thus, the first polynucleotide probe comprises a nucleotide sequences which can hybridize with a sequence of the mRNA that does not occur in other RNA in the sample. Such a sequence is referred to herein as a unique region of the mRNA. Preferably, the first polynucleotide probe does not bind to portions of the target mRNA other than the unique region.

A unique region can be identified by any of several methods. These include: (1) comparing the nucleotide sequence between several potential mRNAs and their target mRNAs; (2) getting a candidate sequence which can be compared with other nucleotide sequences; and (3) computer searching for nucleotide sequences which do not pair with the

target mRNA. Information on various known mRNA sequences can be obtained from several genetic information databases.

5 The first polynucleotide probe is preferably an oligodeoxyribonucleotide (DNA) rather than an oligoribonucleotide (RNA) since DNA is more stable than RNA. The number of nucleotides is not restricted; however, if an oligodeoxyribonucleotide is used as the first probe, a preferred length is 17-40 nucleotides. If a cDNA of target RNA is available, a part or the entire cDNA can be used as the first nucleotide probe. The first nucleotide probe can also be easily manufactured by using a DNA synthesizer or by a commercial manufacturer.

10 In this invention, the first nucleotide probe used is immobilized to an insoluble support. It may be chemically modified for immobilization, such as by adding an amino group onto the 5'-terminal or 3'-terminal of the polynucleotide.

#### Insoluble Support

15 In the practice of the present invention, the first nucleotide probe is immobilized to an insoluble support. The insoluble support can be any of a variety of insoluble materials, such as a plastic plate, membrane filter, microtiter plate or any other insoluble material to which polynucleotides can be attached. The insoluble support is preferably made of a material which can immobilize the first nucleotide probe by covalent bonds. For example, a plastic plate, such as polystyrene, which has a carboxyl group or amino group on the surface thereof can be used. Such plates are available commercially. However, procedures for introducing these residues are also well known in the art. One commercially available plastic plate which has carboxyl group, on its surface is Sumilon microtiter plate MS-3796F made by Sumitomo Bakelite. A plastic plate having an amino group on the surface is the Sumilon microtiter plate MS-3696F.

#### 25 Immobilization of First Polynucleotide

In the practice of the present invention, the first polynucleotide probe is immobilized onto the insoluble support. Various methods of immobilizing polynucleotides to the insoluble support are available, including covalent binding, ionic binding, and the physical absorbance method. However, the covalent binding method is preferred. Thus, in certain  
30 embodiments of the present invention, the polynucleotides are immobilized to microtiter plates which exhibit functional groups, such as carboxyl residues, amine residues, or hydroxyl residues on the surface thereof.

In a preferred procedure for immobilization of the polynucleotide to an insoluble support exhibiting a functional group, the 5'-terminal end of the polynucleotide is covalently

linked to the functional group. Any of a variety of methods for covalent binding of polynucleotides to these functional groups can be used. Examples of preferred well-known methods include the maleimide method and carbodiimide method.

5 The maleimide method, illustrated in Figure 1, involves the reaction between a substance containing a maleimide group and another material containing a sulfhydryl residue (SH). In order to attach the 5' end of a polynucleotide to an immobilized support using the maleimide method, the 5' end of the polynucleotide is reacted with a maleimide compound. A suitable maleimide compound is sulfosuccinimidyl-4-(N-maleimidomethyl) cyclohexane-1-carboxylate (sulfo-SMCC).

10 The SH residue is provided on the support by a reaction between a support having an amino group and succinimidyl-S-acetylthioacetate (SATA), followed by deacetylation using hydroxylamine (NH<sub>2</sub>OH). (Sulfo-SMCC and SATA are readily available from a variety of commercial sources, including the Pierce Company.) The resulting SH group on the support can then react with the maleimide group on the 5' end of the polynucleotide to  
15 form a polynucleotide-immobilized support. One problem we have experienced in the use of the maleimide method is that the SH group on the plate can react not only with an amino group at the 5' end of the polynucleotide, but can also react with primary amino groups on the purine bases, adenine and guanine. In order to assure that the polynucleotides are immobilized at their 5' ends, the amino groups on the purine bases can  
20 be protected by pairing the polynucleotide to a complementary polynucleotide prior to immobilization. After immobilization, the complementary polynucleotide can be removed through denaturation, such as through heating.

The carbodiimide method is illustrated in Figure 2. This method involves a reaction  
25 between an amino group and a material containing a carboxyl residue using a carbodiimide compound. An example of a carbodiimide compound is 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (hereafter called EDC). This reaction can be enhanced with N-hydroxysulfosuccinimide (hereafter called Sulfo-NHS). Both EDC and Sulfo-NHS are available from well known commercial sources, including the Pierce Company.

30 In the practice of a preferred carbodiimide method for attaching polynucleotides to a support, a support having a carboxyl residue attached is used. EDC is activated by reaction with Sulfo-NHS. This activated EDC is reacted with support containing surface-bound carboxyl residues. This can then be reacted with polynucleotides having an amino group at their 5'-terminal ends, resulting in a polynucleotide-immobilized support.

In order to assure that the polynucleotides are immobilized at their 5' end, primary amino groups on adenylyl, guanylyl and cytosyl groups can be protected by hybridizing the nucleotide to a complementary polynucleotide prior to immobilization. After immobilization, the complementary polynucleotide can be removed through denaturation, such as through heating.

We have found that non-specific binding of activated amino or carboxyl residues on insoluble supports can be effectively reduced or eliminated by treating the plates to which polynucleotides have been immobilized with a primary amine compound, preferably glycine.

#### Second Polynucleotide Probe

The second nucleotide probe contains sequences which can also hybridize to target mRNA that is trapped by the first bound nucleotide sequence. Preferably, the second polynucleotide probe hybridizes to a different portion of the target mRNA than does the first polynucleotide probe. In an especially preferred embodiment, the second nucleotide probe consists of nucleotides which are complementary to the polyadenylic acid tail of the target mRNA. The second nucleotide probe is also preferably DNA rather than RNA due to the higher stability of DNA. The number of nucleotides is not limited; however, a polynucleotide of 15-40 nucleotides is preferable. Thus a poly (dT) sequence of 15-40 nucleotides is a preferred second polynucleotide probe. Such a second polynucleotide probe would serve as a universal second probe for virtually any target mRNA from eukaryotic cells. However, if a cDNA to the target RNA is available, a part or all of this cDNA can also function as a second polynucleotide probe. The second nucleotide probe can also be easily manufactured by a DNA synthesizer or commercial manufacturer.

The second polynucleotide probe is preferably labeled in order to easily detect its presence. A variety of chemical substances are available which can label the target substance. For example, a variety of radionuclides can be used, such as the radioisotopes  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ , and  $^{125}\text{I}$ . Enzymes can also be used to label the second polynucleotide probes. Suitable enzymes include alkaline phosphatase, luciferase, and peroxidase. Other labels might be chemical compounds such as biotin, avidin, streptavidin, and digoxigenin. Labels which provide a colorimetric indication or a radionuclide are preferred. In a preferred embodiment, biotin is attached to the nucleotide probe, followed by an avidin, such as streptavidin, conjugated to an enzyme such as alkaline phosphatase. The presence of the enzyme can be detected by various substrates, such as ATTOPHOS which provides a fluorescent marker that can be detected through fluorimetry.

In certain embodiments, part or all of the labeling compound is substituted with part or all of the second nucleotide probe. Techniques for substituting portions of polynucleotides are well known for a variety of labels and kits are commercially available for many of these. For example, one can use a DNA 3'-end Labeling Kit or oligodeoxythymidylic acid to label the probe.

Nucleic acid can also be "labeled" by staining with a nucleic acid stain. Thus, where a relatively large amount of nucleic acid is present, ethidium bromid (EtBr) can be used to identify the presence of nucleic acid. This is useful within the context of the present invention, for example, if mRNA has hybridized to a short immobilized anti-sense probe then substantially more nucleic acid will be immobilized than if no mRNA had hybridized. However, more sensitive stains are more preferable. Such more sensitive stains include the various cyanine nucleic acid stains, such as POPO, BOBO, YOYO and TOTO available from Moecular Probes (California). These stains are described, e.g., in *Science*, **257**:885 (1992). Particularly preferred stains for use in the context of the present invention include the shorter wavelength forms, TOTO-1 and YOYO-1, still more preferably YOYO-1. As little as four picograms of stained DNA can be detected by visible fluorescence upon stimulation with a transilluminator or hand-held UV lamp. Thus, these stains provide a particularly easy and sensitive method of identifying the presence of nucleic acids.

We tested the ability of YOYO-1 to stain oligonucleotides immobilized to wells as discussed herein. We first immobilized various known amounts of oligonucleotides to wells and washed to remove non-immobilized oligonucleotides. We then added 1:1000 dilution of YOYO-1 in water. We then used a fluorimeter directly without washing. The relation between pmoles of oligonucleotide is shown in Figure 15 with open circles. We also incubated the TOTO-1 stained immobilized oligonucleotides for ten minutes, washed and added water, followed by use of the fluorimeter. The washed results are shown in dark circles in Figure 15. It can be seen that washing does not significantly effect the amount of staining and that the amount of staining is related to the amount of oligonucleotide.

We also tested the time course of YOYO-1 staining over the course of one hour. We included a control in which no oligonucleotide was immobilized to the plate. The oligonucleotide-immobilized plate is shown in open circles in Figure 16 and the control in dark circles in that Figure. It can be seen from Figure 16 that after an initial spike, relatively constant staining is found.

We further tested the dose response of constant amounts of oligonucleotides with oligonucleotides immobilized in wells (open circles in Figure 17) and control wells with no



oligonucleotide (closed circles in Figure 17). The difference between the immobilized and non-immobilized is shown in Figure 17 as triangles. It can be seen that a sharp increase in fluorescence occurs between  $10^{-4}$  and  $10^{-3}$  dilution.

We also tested the reproducibility of staining of both oligonucleotide-immobilized wells and non-immobilized wells. We repeated the experiment five times and graphically depicted the data for both oligonucleotide immobilized (+) and non-immobilized (-) wells in Figure 18. It can be seen that the data was substantially similar for each experiment.

Thus, the data depicted in Figures 15-18 shows that use of YOYO-1 as a staining agent provides a reliable indication of the amount of oligonucleotide present.

#### Application of Sample

In the preferred methods of the present invention, a sample containing mRNA is applied to the insoluble support to which the first polynucleotide probe has been bound. Any target mRNA found in the sample is allowed to hybridize with the first polynucleotide probe. This can be accomplished by incubation at temperature dependent on a variety of factors, as is well known to those with ordinary skill in the art. These factors include the length of complementary nucleotide sequences, the ratio of Guanine or Cytosine (GC contents) within the entire base content in the complementary nucleotide sequences, NaCl concentration in buffer, the number of bases which mismatch in the complementary nucleotide sequence, and the type of nucleotide. In a preferred form of this invention, the following equation can be used to calculate the preferred incubation temperature:

$$T_{inc.} = 16.6 \times \log (M) + 0.41 (GC) + 81.5 - 675/n - 15 (^\circ C).$$

In the equation shown above, M is the NaCl concentration (M) in solution, GC represents GC contents (%), n represents the length of nucleotide sequences (the number of nucleotides).

The incubation time can also be determined according to methods described in the Molecular Cloning manual.

The time for incubation is preferably from 1 hour to overnight, and the sample should preferably be gently swung during incubation. Incubation is preferably performed in an appropriate buffer solution. The same buffer used to hybridize RNA and DNA in the Northern Blot or the Dot Blot Methods can be used. The buffer is preferably prepared in a way so as not to contaminate it with RNase. If any RNase contamination is present, the activation thereof should be controlled to be as low as possible. RNase-free buffers are commercially available, such as within mRNA purification kits or lysis buffer as shown before.

In order to eliminate RNase activity from water used in the methods of this invention, the water is preferably treated with Diethylpyrocarbonate (DEPC). The preferred DEPC treatment involves addition of 0.1% DEPC to the water, followed by storage overnight at 37°C and sterilization in an autoclave.

5     **Washing Procedures**

Preferably, after incubation, unbound components of the sample are washed from the insoluble support. Appropriate solutions for washing include the lysis buffer as shown before or the buffer included in the mRNA purification kit.

10     **Application of Second Polynucleotide Probe**

The second polynucleotide probe is applied to the insoluble support after any washing of unbound components of the sample, under conditions as described above in connection with the hybridization of mRNA in the sample with the first nucleotide probe. This is preferably followed by washing, as described above.

15     **Quantitation**

The amount of mRNA in the sample is quantitated by measuring the amount of second polynucleotide probe bound to the insoluble support after application thereto in accordance with the present invention. In this regard, a physical or chemical quantity or activity of the label on the second polynucleotide can be measured by several techniques, including optical density, emitted-light intensity, or radiation. The label itself can provide this indication or can require other compounds which bind thereto or catalyze the label. Other mechanisms for detecting label include the use of compounds that can chemically react against the label, detection of colored label, detection of light emission, detection of radiation or of catalytic ability.

20     In a preferred measurement technique, the label on the second polynucleotide is biotin. The presence of this label can be detected by reaction with the enzymes Peroxidase or Alkaline phosphatase. These enzymes can be specifically directed to biotin by conjugation with avidin or streptavidin. The presence of the enzymes is then detected by addition of an appropriate substrate to provide a detectable color-developing or light-emitting reaction. The alkaline phosphatase labeled streptavidin can be readily obtained from the commercial market. A preferred substrate for alkaline phosphatase is adamantyl-30     1,2-dioxetane phosphate (AMPPD). Upon reaction with the alkaline phosphatase, AMPPD will emit light at a wavelength of 447 nm. This light can be detected in accordance with techniques known in the art.

In the reaction between alkaline phosphatase and AMPPD, an enhancer such as 5-N-tetradecanoyl-amino-fluorescein can be added. 5-N-tetradecanoyl-amino-fluorescein has the ability to convert light of 477 nm wavelength to 530 nm wavelength which is more readily detectable.

5 Other preferred labels include an antigen, such as digoxigenin or an antibody. An antigen can be detected by its ability to bind to an antibody directed thereto. Such antibodies, or an antibody directly used to label the second polynucleotide probe, can be detected by their ability to bind protein A. The antibody itself can be labeled directly with a radionuclide, such as  $^{125}\text{I}$ , or can be labeled by binding thereto of protein A labeled with  
10 the radionuclide. The radionuclide can then be detected in accordance with techniques well known in the art.

Well-known techniques for the detection of label include the detection of label which appears by a color-developing reaction by a spectrophotometer. Other such techniques include the detection of a light emitting reaction using X-ray film or an instant camera, and  
15 the detection of radiation using a radiation counter.

The emission reactions are recorded by X-ray film or instant camera film in the dark room. The X-ray film which is exposed by emission reactions are recorded as a blot, so this shading of the blot can be measured by a densitometer. If one uses an instant camera such as a Polaroid, the picture is read by a scanner to decide the location of the blot on the  
20 computer, measuring the shading of the blot using graphic analysis software.

To measure the amount of target mRNA in an unknown sample solution, an arithmetic plot for showing the relationship between a standard solution of target mRNA and relative concentration in a blot can be used. The standard solution for target mRNA need not be purified target mRNA. Thus, the standards can also be a mixture in which the  
25 amount of target mRNA is known.

#### Schematic Example

A schematic representation of an exemplary method in accordance with the present invention is shown in Figure 3. The target mRNA 1 is shown having a polyadenylic acid tail  
2. The first nucleotide probe 3 is immobilized to an insoluble support 4, then mixed the  
30 target mRNA 1. The target mRNA 1 hybridizes with the first nucleotide probe 3, thereby immobilizing the target mRNA to the insoluble support. Then, a second nucleotide probe 5 which bears a label 6 hybridizes with a portion of the target mRNA other than the portion hybridized with the first nucleotide probe. As shown, the second nucleotide probe 5

comprises poly (dT) and hybridizes with the polyadenylic acid tail 2 of the target mRNA 1. The resulting complex 7 can then be identified by detecting the presence of the label 6.

**Time-Savings of Invention**

5        Advantageously, the present invention provides a method of measuring mRNA which can be performed in far less time than previously known methods. Table 1 hereinbelow compares the time needed to measure DNA using the method of this invention and a Northern blot method.

**TABLE 1**  
**Comparison For Measurement Of mRNA**

5	Present Invention		Northern Blot Method	
		Time (hours)		Time (hours)
10	- Lysis of Cells	1	- Purification of mRNA	48
	- Hybridization of the first nucleotide probe to mRNA	1	- Separation of mRNA by electrophoresis with Agarose Gel	1
			- denaturing of mRNA moving the mRNA to membrane	1
15			- immobilization of mRNA to membrane	2
	- hybridization of the second nucleotide probe to mRNA	1	- hybridization of the mRNA and probe	12
20	- washing and blocking	0.2	- washing and blocking	4
	- detection	1	- detection	1
Total		about 5 hours		about 70 hours

As shown in Table 1, the methods of this invention generally take 5 hours or less to complete, while the Northern blot method generally requires about 70 hours. These times do not include the labeling time (about 12 hours) of the nucleotide probes nor the time for immobilization of the first nucleotide probe because these constructs can be prepared prior to the measurement.

Particular examples of target mRNA that can be measured to obtain clinically useful information are provided below. The examples included herein are intended to illustrate certain preferred embodiments of the present invention. These examples are not intended to limit the invention in any manner.

#### **$\beta$ Receptor**

A potential target for using mRNA measurement to determine a diseased state utilizes the mRNA for the  $\beta$  receptor. The  $\beta$  receptor is a protein located in human nerve tissue. In particular, abnormalities in the  $\beta_2$  receptor has been found to be closely correlated with asthma. Thus, measuring the mRNA for  $\beta_2$  receptor can be used to determine the pathophysiology of asthma patients, and could also be used to assess the effectiveness of anti-asthma agents.

**EXAMPLE 1****Measurement of Human  $\beta_2$  Receptor mRNA  
(Preparation of Calibration Line)****(1) Preparation of the First Nucleotide Probe**

5 An oligodeoxyribonucleotide probe containing the nucleotide sequence shown below with an amino group attached to its 5' end was synthesized by Genosys, Inc. (TX). In this sequence, A, C, G and T indicate adenine, cytosine, guanine, and thymine, respectively. Based on computer analysis (DNASIS, Hitachi Software Engineering America, Ltd., CA), this sequence was determined to be homologous to the human  $\beta_2$  adrenergic receptor mRNA.

10 5'-NH<sub>2</sub>-ATG CTG GCC GTG ACG CAC AGC A-3' (Seq ID NO: 6)

**(2) Immobilization of the First Nucleotide Probe Onto Insoluble Supports**

Both EDC and sulfo-NHS (Pierce, IL) were dissolved in DEPC-treated water at concentrations of 20 mM and 10mM, respectively. EDC/Sulfo-NHS solution was then prepared by mixing equal volumes of both EDC and sulfo-NHS. The first nucleotide probe was dissolved in DEPC-treated water at a concentration of 1 $\mu$ g/ $\mu$ l and then mixed with the EDC/Sulfo-NHS solution in the ration 1:25 (Vol:Vol).

50 $\mu$ l of this probe solution was added to each well of a microtiter plate (MS-3796F, Sumitomo Bakelite, JAPAN), which were known to express carboxyl groups on their surface. After incubation at room temperature overnight, the reaction solution was removed by aspiration.

**(3) Preparation of mRNA**

DNA plasmid pUC18 containing the human  $\beta_2$  adrenergic receptor cDNA (pTF, American Type Culture Collection) was digested with *Pst*I and *Eco*RI, and mixed with the pBC phagemid (Stratagene, CA) which had been predigested with the same restriction enzymes. The resultant plasmid was linearized by *Pst*I, extracted once with phenol-chloroform, precipitated in ethanol, and used to transcribe the human  $\beta_2$  receptor mRNA using the Riboprobe system and T7 RNA polymerase (Promega, WI). The transcribed human  $\beta_2$  receptor mRNA was treated with RNase-free DNase to digest the remaining template cDNA, then purified with one cycle of phenol extraction and ethanol precipitation. The resultant mRNA showed up as a single band on agarose (FMC bioproducts, ME) gel electrophoresis. Messenger RNA concentration was measured by a spectrophotometer at 260 and 280 nm against standard mRNA solutions prepared at concentrations of 25 pg, 250 pg, 2.5 ng, 25 ng and 250 ng, respectively.

(4) **Hybridization of the First Nucleotide Probe to mRNA**

In order to remove RNase from the well, the first nucleotide probe-immobilized wells were treated with 250  $\mu$ l of lysis buffer containing 0.5 M NaCl at 45°C for one hour. The buffer was removed from individual wells by aspiration, and 50  $\mu$ l of lysis buffer containing varying concentrations of standard human  $\beta_2$  receptor mRNA was added into each well. These solutions were incubated at 59°C for one hour to allow hybridization.

(5) **Preparation of the Second Nucleotide Probe**

The second nucleotide probe was a 15-mer oligo(dT) labeled with biotin at the 5' end.

(6) **Hybridization of the Labeled Second Nucleotide Probe to the Poly(A) tail of mRNA**

After hybridization of the first nucleotide probe and mRNA described in step (4), the hybridization solution was removed by aspiration and the individual wells were washed with 250  $\mu$ l of lysis buffer. Fifty  $\mu$ l of lysis buffer containing 0.5 M NaCl and 1  $\mu$ l of the biotinylated second probe solution (35 pmol/ $\mu$ l) was added into each well, and incubated at room temperature for an additional one hour.

(7) **Measurement of Chemical Activities of the Labeled Second Nucleotide Probe**

Following the hybridization described in step (6), the hybridization solution was removed by aspiration and wells were washed once with 250  $\mu$ l of lysis buffer. A blocking buffer consisting of (0.05%(w/v) Tween 20, 500 mM NaCl, 100 mM Tris-HCl, pH 7.5) was added into each well and incubated at room temperature for five minutes to reduced nonspecific binding. Reaction solutions were then removed by aspiration, 50  $\mu$ l of blocking solution containing 1/5000 vol. of alkaline phosphatase-conjugated streptavidin solution (Clontech, CA) was added into each well, and incubated at room temperature for 30 minutes. Each well was then washed with 250  $\mu$ l of washing buffer containing 500 mM NaCl, 100 mM Tris-HCl, pH 7.5, and 50  $\mu$ l of Lumiphos 530 solution (Boehringer-Mannheim, IN) whole plates were placed on X-ray film (XAR5, Kodak, NY) in specific holders (MYC, Kodak, NY). X-ray film was exposed to a generated chemiluminescent light at room temperature for one hour, and then developed using a film processor (QX-400, Konica, Japan). Relative intensities of individual wells on X-ray films were quantified by densitometry (model 620, Bio-Rad, CA).

(8) **Results**

The results of Example 1 are shown in Figure 8. The inset (A) shows results of an X-ray film exposed to chemiluminescent light during the detection of mRNAs by the invented method; (B) shows a densitographic representation of the same image as in (A):

(C) exemplifies a calibration line of the mRNA in which the X-axis represents concentrations of mRNA and the Y-axis represents the relative intensity of chemiluminescence in an individual well. Wells 1, 2, 3, 4 and 5 of Figure 8 correspond to the amounts of tested mRNA at 25 pg, 250 pg, 2.5 ng, 25 ng and 250 ng, respectively. As shown in (A), mRNAs were detected in amounts over 25 pg by the invented method. The calibration line shown in (C) shows the high linearity of signal when increasing concentrations from 25 pg to 250 ng of mRNA.

#### COMPARISON 1

##### Quantification of mRNA by Northern Blot

###### 10 (1) Agarose Gel Electrophoresis of mRNA

Ten mM of sodium phosphate buffer, pH 7.0 containing 0.1% DEPC was heated to 37°C overnight and autoclaved. A 1% agarose solution was prepared using this 10 mM DEPC-treated phosphate buffer. The solution was chilled to ca. 50°C, and iodine acetate was added to 0.2% (W/V). A gel slab was formed (6.5 cm x 10 cm) with combs and placed at room temperature for one hour to allow hardening. Five µl of DEPC-treated water containing 250ng, 25 ng, 2.5 ng, 250 pg or 25 pg of mRNA described in example 1.(3), were mixed with 7 µl of loading buffer (125 µl DMSO, 42 µl 40% glyoxal, 3 µl 1 M sodium phosphate buffer, pH 7.0), and incubated at 50°C for one hour. Samples were applied to the gel, and electrophoresed in a solution of DEPC-treated 10 mM phosphate buffer at constant 6V/cm for 15 minutes without circulation. An additional one hour of electrophoresis was performed with circulation.

Following electrophoresis, gels were immersed in 50mM sodium hydroxide containing 0.5 µg/ml ethidium bromide (Sigma, MO) for 20 minutes, and 0.1 M Tris-HCl buffer, pH 7.5 for 40 minutes. Gels were then transferred into fresh 0.1 M Tris-HCl buffer and photographed under UV light.



(2) Northern Transfer of mRNA Onto Membranes

Filter paper (3MM Chr, Whatman, NJ) and a nylon membrane (MagnaGraph, Micron Separations, MA) were immersed in sterile water for 20 minutes and then placed in 10x SSPE (1.5 M NaCl, 115 mM sodium phosphate, 11.5 mM EDTA in water, pH 7.4 adjusted with 10 N NaOH). A blotting sponge (Stratagen, CA) was also immersed in 10x SSPE. The 3MM filter, nylon membrane, agarose gel, and sponge were assembled from the bottom in a Posi-blot Pressure blotter (Stratagen, CA), to cause the mRNA to be transferred onto the nylon membrane by positive pressure (835 mmHg for 45 minutes). The nylon membrane was dried, and the mRNA was permanently immobilized onto the nylon membrane by UV light cross linkage at 120 mJoules (Stratalinker 1800, Stratagen, CA).

(3) Preparation of Biotinylated cDNA Probe

DNA probes were prepared from  $\beta_2$  receptor cDNA by using a DNA labeling kit (Boehringer-Mannheim, IN) and biotin-labeled deoxyuridylic acid (Clontech, CA). In brief, a  $\beta_2$  receptor cDNA-containing plasmid, pTF was digested with *Pst*I and *Eco*RI. The resultant plasmid and cDNA were separated by agarose gel electrophoresis. A band of  $\beta_2$  receptor cDNA was cut out and removed from the gel using the Gene clean kit (Bio 101, CA). 0.1  $\mu$ g of above cDNA dissolved in 5  $\mu$ l of water was incubated at 95°C for 10 minutes, and then incubated with 2  $\mu$ l of hexanucleotide mixture, 2  $\mu$ l of dNTP and 1  $\mu$ l of Klenow enzyme (DNA labeling kit, Boehringer-Mannheim, IN) at 37°C for one hour. Following incubation, the reaction was terminated by addition of 2  $\mu$ l of 0.2 M EDTA, pH 8.0, 2.5  $\mu$ l of 4.0 M LiCl and 75  $\mu$ l of 100% ethanol, and cooled at -70°C for 30 minutes to precipitate cDNA. After centrifugation, the pellet was washed once with 70% ethanol and then dissolved in 50  $\mu$ l of water.

(4) Hybridization of Biotinylated cDNA Probe to mRNA

The mRNA containing membranes were incubated in hybridization buffer containing 0.5 M NaCl (ECL, Amersham, IL) at 45°C for one hour. Biotinylated cDNA probe described above were first heat denatured at 95°C for 10 minutes, immediately cooled on ice, and then added to the hybridization buffer. Hybridization was continued at 40°C overnight.

(5) Chemical Activity Measurements of Labeled cDNA Probe

After hybridization, membranes were washed twice in 0.5x SSPE solution containing 36% (W/V) urea, 0.4% (W/V) SDS and 0.5 M NaCl at 65°C for 15 minutes each, followed by washing twice in 2x SSPE with 0.5 M NaCl at room temperature for five minutes each. Membranes were then incubated with 100 mM Tris-HCl buffer, pH 7.0, containing 2%

blocking reagent (Boehringer-Mannheim, IN) and 150 mM NaCl at room temperature for three hours. 1/5000 vol. of alkaline phosphatase-conjugated streptavidin (Clontech, CA) was added to the above solution and incubated at room temperature 30 minutes. Membranes were washed twice with 100 mM Tris-HCl, pH 7.0, containing 150 mM NaCl at room temperature for 30 minutes each followed by one washing with alkaline phosphate buffer (100 mM NaCl, 5 mM MgCl<sub>2</sub> in Tris-HCl buffer, pH 9.5) at room temperature for 2 minutes. Membranes were then soaked in Lumi-Phos 530 (Boehringer-Mannheim, IN), and quickly placed between two sheets of transparent plastic film. Chemiluminescent membrane signals were exposed to X-ray film (XAR5, Kodak, NY) in specific holders (MYC, Kodak, NY) at room temperature for one hour.

(6) **Results**

The Northern blot results of Comparison 1 are shown in Figure 8, in which the chemiluminescent signals of mRNA were exposed to an X-ray film. Lanes 1, 2, 3, 4, and 5 of Figure 8 correspond to 25 pg, 250 pg, 2.5 ng, 25 ng and 250ng of applied mRNA, respectively. Because this cDNA probe is approximately 100 fold larger than the oligonucleotide probe used in Example 1, the amount of incorporated biotin is also 100 fold more than the oligonucleotide. This result indicates that the cDNA probes used in Comparison 1 are approximately 100 fold more sensitive than the oligonucleotide probe. However, as seen in Figure 8, a conventional Northern blot can only detect samples containing greater than 25 ng of mRNA. This is far less sensitive than the invented method as exemplified by Figure 8. Also, the invention as described by Figure 8 used less sensitive oligonucleotide probes when compared to the cDNA probes used on the Northern Blot.

**EXAMPLE 2**

**Detection of Human  $\beta_2$  Receptor mRNA From Unknown**

**Samples And Evaluation of Reproducibility**

(1) **Preparation of the First Nucleotide Probe and Its Immobilization Onto the Insoluble Support**

The first nucleotide probe was prepared and immobilized onto an insoluble support as described in Example 1.

(2) **Preparation of Standard mRNA**

mRNA was prepared as described in Example 1.

(3) **Preparation of Cell Lysates**

5 Human HL60 promyelocytic leukemia cells were collected by centrifugation, washed with PBS (8 g NaCl, 0.2 g KCl, 1.44 g  $\text{Na}_2\text{HP}_4$  and 0.24 g  $\text{KH}_2\text{PO}_4$  in 1 liter of water, pH 7.4). Cells were then resuspended in lysis buffer at a concentration of  $1 \times 10^7$  cells/ml. The solution was repeatedly passed through a sterile plastic syringe fitted with a 21 gauge needle and incubated in a slow-shaking waterbath at 45°C for one hour.

(4) **Hybridization and Chemiluminescent Detection**

10 As described in Example 1, hybridization of the first nucleotide probe, preparation of the second nucleotide probe, hybridization of the labeled second nucleotide probe to the poly(A) tail of mRNA, and chemical activity measurement of the labeled second nucleotide probe was carried out.

(5) **Preparation of Calibration Line**

15 As described in Example 1, a calibration line of known  $\beta_2$  receptor mRNA concentrations was created (Figure 9). The calibration line expresses a linear correlation between mRNA concentrations mapped to a logarithmic scale and the relative chemiluminescent intensity read from X-ray film.

(6) **Evaluation of Reproducibility**

20 Concentrations of  $\beta_2$  receptor specific mRNA from HL60 cell lysates were determined three times. In Table 2, each concentration of  $\beta_2$  receptor specific mRNA was determined from the calibration line described in section (5) above. The mean and standard error (S.E.) of data were calculated. Cell numbers were counted by a phase hemacytometer (American Scientific Products, IL). As shown in Table 2, three separate assays exhibited  
25 similar mRNA values at a range of 0.96 to 1.12 fg/cell. This suggests that the invented method provides reproducible method of quantifying specific mRNA levels.

TABLE 2

Concentrations of  $\beta_2$  Receptor-Specific mRNA in HL60 Cells

No. of assay	mRNA Concentration [fg/cell]
1	1.01
2	0.96
3	1.12
mean $\pm$ S.E.	1.03 $\pm$ 0.04

G Proteins

Cell surface receptors for hormones and neurotransmitters are known to be coupled to intracellular heterotrimeric GTP-binding proteins (G proteins) composed of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits. Once receptors are activated by specific ligands, receptor-coupled G proteins transduce signals to intracellular secondary effector systems, such as adenylyl cyclase, phospholipase C, and ion channels.

Recently, the molecular mechanisms of the interaction between receptors and G proteins have become less clear, because extensive molecular cloning has identified more subclasses of G protein  $\alpha$  subunits than there are known cellular functions of G proteins. Furthermore, it has been found that a single receptor is also capable of coupling to multiple G proteins. For example, histamine  $H_2$  receptors in human HL-60 cells were coupled to two different G proteins, which activated adenylyl cyclase and phospholipase C, respectively. It has also been reported that GABA-B receptors were capable of coupling to  $G_0$ ,  $G_0^*$  and  $G_{i-1}$  proteins, but not  $G_{i-3}$  proteins. Moreover, CHO cells have been shown to express multiple G proteins, which activate the same phospholipase C pathways but are coupled selectively to different receptors. Some receptors may alter their G-protein subclass association depending on the state of differentiation of the cell or on the particular phase of the cell cycle. One example would be the mRNA coding for G protein. G protein binds to and controls the function of various receptors in the cell membrane. It catalyzes the intracellular reactions of cAMP and calcium ion free reaction systems. These molecules are used in various cellular transmitter systems.

G proteins are believed to be involved in causing various disease states. For example, a genetic deficiency of  $G_s$  proteins is the molecular basis of hereditary

osteodystrophy. Pituitary tumors in acromegalic patients have been shown to contain mutant Gs proteins. G proteins are also involved in invasive and metastatic melanoma cells. Rat models of streptozotocin-induced experimental diabetes suggest that the levels of mRNA for various subclasses of G $\alpha$  proteins are significantly altered from normal control rats. Furthermore, cellular functions of pertussis toxin-sensitive G proteins were shown to be significantly impaired in atherosclerotic porcine coronary arteries, while G protein function in leukocytes of patients with mania was hyperfunctional. Thus, those diseases which are known to be related to a malfunction of the G protein include atherosclerosis, bipolar disorder, diabetes mellitus, melanoma, and pituitary tumors.

Currently available immunological detection methods (Western blots) and mRNA detection methods (Northern blots) are not sensitive and require a lot of cellular material, making it difficult to study the role of G proteins in disease. It is not easy to ascertain the quantity of expressed G protein. If the mRNA for the G protein was measured, it could be used to recognize the pathophysiology of patients having one of the foregoing ailments. The effects of agents designed to arrest or control these conditions could also be measured using this technique. G protein consists of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits. The  $\alpha$  subunit is known to have several allelic variations. Thus, by measuring each variation of G protein mRNA, the amount of expressed protein for each variation can be estimated.

### **EXAMPLE 3**

#### **Quantification of Human G Protein $\alpha$ Subunit-Specific mRNA**

##### **(1) First Nucleotide Probe Preparation**

The first nucleotide probes used against Gs protein-specific mRNA and Gi-2 protein-specific mRNA were oligodeoxyribonucleotides. These nucleotide sequences, shown below, contain an amino group on their 5' end and were synthesized by Genosys (Texas).

The probe sequence used for quantifying Gs mRNA is shown below (Seq ID No. 1):

5'-NH<sub>2</sub>-TTC ATC CTC CCA CAG AGC CTT G-3'

The sequence of probe used for quantifying Gi-2 mRNA is shown below (Seq ID No. 2):

5'-NH<sub>2</sub>-ATG GTC AGC CCA GAG CCT CCG G-3'

(2) **Immobilization of the Second Nucleotide Probe Onto Insoluble Supports**

The first nucleotide probes were dissolved in DEPC-treated water at  $1\mu\text{g}/\mu\text{l}$ . Probes were then mixed with the EDC/Sulfo-NHS solution described in Example 1 Paragraph (2) in a ratio of 1:25 (Vol:Vol). A  $50\mu\text{l}$  sample was added to each well of a Sumilon microtiter plate (MS-3796F, Sumitomo BAKelite, JAPAN) and incubated overnight at room temperature. Following incubation, the reaction solution was removed with an aspirator.

(3) **Preparation of Standard mRNA**

A plasmid vector pGEM2 containing rat Gs protein cDNA (provided by Dr. R. R. Reed, Johns Hopkins University, U.S.A.) was linearized by digestion with *NheI* (Promega, WI). Gs protein-specific mRNA was prepared by the method described in Example 1 Paragraph (3) using T7 RNA polymerase (Promega, WI). The resultant mRNA concentrations were determined by spectrophotometry and subsequently diluted with DEPC-treated water as described in Example 1 Paragraph (3).

(4) **Preparation of Cell Lysates**

Human B lymphocyte-derived IM9 cells were centrifuged, washed with PBS, and resuspended in lysis buffer at  $1 \times 10^7$  cells/ml as described in Example 2. Cells were then lysed by repeated passage through a sterile plastic syringe fitted with a 21 gauge needle, and then incubated in a slow-shaking waterbath at  $45^\circ\text{C}$  for one hour.

(5) **Hybridization of the First Nucleotide Probe to mRNA**

In order to remove RNase contamination in the microtiter well,  $250\mu\text{l}$  of lysis buffer was added to each well at  $45^\circ\text{C}$  for one hour as described in Example 1. The buffer was aspirated, and  $50\mu\text{l}$  of lysis buffer containing 0.5 M NaCl and various concentrations of human Gs protein mRNA was added into each well. Each sample was incubated at  $51^\circ\text{C}$  for one hour to allow hybridization as described in Example 1.

(6) **Hybridization of the Labeled Second Nucleotide Probe to a Poly(A) Tail of mRNA, and Measurement of Chemical Activities of the Labeled Second Nucleotide Probe**

After hybridization of the first nucleotide probe, the hybridization solution was aspirated off and individual wells were washed once with 250  $\mu$ l of lysis buffer. Fifty  $\mu$ l of lysis buffer containing 0.5 M NaCl and 1  $\mu$ l of the biotinylated second nucleotide probe solution (35 pmol/ $\mu$ l) was added into each well, and incubated at room temperature for one hour to allow hybridization.

Chemical activities of the labeled second nucleotide probe was measured as described in Example 1.

(7) **Preparation of Calibration Line**

The calibration line of standard Gs protein-specific mRNA is shown in Figure 10. The calibration line expresses a relationship between the common logarithm of mRNA concentrations and the relative intensity of chemiluminescent signals exposed to X-ray film.

(8) **Evaluation of Reproducibility**

Concentrations of Gs protein and Gi-2 protein specific mRNAs from IM9 cell lysates were measured six times to test reproducibility of the present invention. The methods used for Gi-2 were similar to that described above for Gs, with the incubation temperature used for hybridization of Gi-2 mRNA being 59°C. Table 3 lists the mRNA concentrations as obtained from the calibration line, and, the mean, and the standard error. Cell numbers from each sample were counted with a phase hemacytometer (American Scientific Products, IL). As listed in Table 3, all six assays reported similar mRNA concentrations for both Gs and Gi-2 mRNA. This suggests that the current invention provides a reliable, reproducible means of measuring specific mRNA concentrations.

TABLE 3

No. of Assay	mRNA Concentration [fg/cell]	
	Gs	Gi-2
1	24.8	4.6
2	10.6	8.8
3	18.8	7.2
4	35.5	6.8
5	13.6	11.5
6	33.0	13.7
mean $\pm$ S.E.	22.7 $\pm$ 3.8	8.77 $\pm$ 1.24

**EXAMPLE 4**

Simultaneous Quantification of Various mRNAs  
Derived From Cell Lines by the Invented Method  
in Conjunction with Reverse PCR Method

(1) Simultaneous Detection of Gs and Gi-2 Protein-Specific mRNAs in Various Cell Lysates by the Invented Method

Using the method described in Example 3, specific mRNAs for both Gs and Gi-2 protein from various cell lysates, such as Jurkat human T-cells, IM9 human B cell, U937 human monocytes and HL60 human granulocytic cells were analyzed. Chemical activities of the labeled second probe were exposed to Polaroid film (667, Polaroid, MA) in a specific holder (model 901, Tropix, CA) at room temperature for 5 seconds, instead of exposing to X-ray films as described in Example 3.

(2) Simultaneous Detection of Gs and Gi-2 Protein-Specific mRNAs by the Conventional Reverse PCR Method

(a) cDNA Preparation and Amplification by PCR

mRNA was purified using the FastTrack kit (Invitrogen, CA) directly from various cell lysates as described in Example 3 Paragraph (4). Samples were dissolved in 10  $\mu$ l DEPC-treated water, and cDNA was synthesized using the Superscript kit (Gibco/BRL Life



Technologies, MD). Resultant cDNA was extracted once with phenol, precipitated in ethanol, and dissolved in 10 µl of DEPC-treated water.

One µl of solution containing cDNA was mixed with 1 µl each of 10 mM dATP, dCTP, dGTP and dTTP (Pharmacia LKB Biotechnology, NJ), 1 µl of 25 mM MgCl<sub>2</sub>, 1 µl each of G protein-specific PCR primers, 0.5 µl of *Taq* polymarese (Promega, WI), and 5 µl of PCR buffer (Promega) to a final volume of 50 µl. Sense and antisense G protein-specific PCR primers are exhibited below

AGCACCATTGTGAAGCAGATGA (sense) (Seq ID No. 3)

CTCTGGCCTCCCACATCAAACA (5' to 3' antisense) (Seq ID No. 4)

These sequences are known to be highly conserved among various G proteins, with this pair of primers being known to amplify Gs, Gi-1, Gi-2, Gi-3 and Go protein specific cDNA from data gathered in preliminary experiments. Followed by pre-denaturation at 95°C for 10 minutes, PCR was carried out in a DNA thermal cycler (Perkin-Elmer Cetus, CT) with 30 cycles of annealing at 55°C for 1.5 minutes, extension at 72°C for 4 minutes, and denaturation at 95°C for 1.5 minutes, respectively.

(b) Agarose gel electrophoresis of amplified cDNA

A 1.2% agarose gel containing 0.00005% ethidium bromide, 0.001 M EDTA and Tris-acetate, pH 8.0 was prepared. Ten µl of each test solution was mixed with 2 µl of gel-loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol and 15% Ficoll (Type 400, Pharmacia LKB Biotechnology, NJ; in water) and heated at 65°C for five minutes prior to loading. Electrophoresis was performed at 6 V/cm for one hour. After electrophoresis, the cDNA bands were examined under UV light illumination.

The gel was then immersed in 0.25 M HCl for 30 minutes (depurination), 0.5 M NaOH containing 1.5 M NaCl for one hour (denaturation), followed by 1 M Tris-HCl, pH 7.5 containing 1.5 M NaCl for one hour (neutralization).

(c) Transfer of Amplified cDNAs onto Membranes (Southern Blot)

Amplified cDNAs were transferred onto nylon membranes and immobilized by UV light as described in Comparison 1(2).

(d) Preparation of Biotin-Labeled cDNA Probes

Ten ng each of Gs and Gi-2 cDNA clones were amplified by PCR as described in (2)(a) above, with the modification that 20% of dTTP was replaced with biotin-labeled dUTP (Clontech). In preliminary Southern blot analysis, these biotin-labeled PCR products have been shown to specifically hybridize to Gs and Gi-2, respectively.

(e) Hybridization of Labeled cDNA Probes to Southern Membranes

As shown in Comparison Example 1 Paragraph 4, biotin-labeled PCR products were heat-denatured at 95°C for 10 minutes, then hybridized to Southern membranes, on which PCR-amplified cDNAs from cell lysates were immobilized.

(f) Chemical Activity Measurement of Labeled cDNA Probes

Chemical activity of labeled cDNA probes was measured as described in Comparison 1 Paragraph (5).

(3) Results

The results of Example 4(1) are shown in Figure 11. Polaroid film was exposed to chemiluminescent light derived from an oligo-(dT) probe by the invented method, in which Gi-2 and Gs protein mRNA from various cells were trapped onto the first oligonucleotide-immobilized plastic plate. Phosphate-buffered saline (PBS) was used as a control. As shown in Figure 11, mRNA was detected from all cell lines.

The results of Example 4(2) are shown in Figure 12, in which Gi-2 and Gs cDNA probes were used on a Southern Blot. PCR-amplified G protein specific DNA from four different cell lines were immobilized onto a membrane. Figure 12 is a copy of the X-ray film detailing the chemiluminescent results of this probe. As shown in Figure 12, mRNAs are detected in all cell lines. Therefore, the invented method provides a high enough reliability to be used for quantification of mRNA.

EXAMPLE 5

Methods for Amplifying G Proteins With PCR Primers

(1) Materials

The cDNAs of rat G protein  $\alpha$  subunits ( $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_s$  and  $G_o$ ) were provided by Dr. R.R. Reed (Johns Hopkins Univ., MD).  $\lambda$ ZAP libraries of rat pituitary and intestine were provided by Dr. D.G. Payan (Univ. Calif. San Francisco). Kirsten murine sarcoma virus transformed rat kidney cells (KNRK), human IM9 B-lymphocytes and human Jurkat T-lymphocytes were obtained from American Type Tissue Culture Collection, Rockville, MD). Cell culture media, Superscript (Gibco/BRL, Gaithersburg, MD), reagents for PCR (Promega, Madison, WI), ECL (Amersham, Arlington Height, IL), Genius, Lumi-Phos 530 (Boehringer-Mannheim, Indianapolis, IN), FastTrack (Invitrogen San Diego, CA),  $\lambda$  gt10 library of human HL-60 cells, biotin-dUTP, alkaline phosphatase-conjugated streptavidine (Clontech, Palo Alto, CA), dNTP (Pharmacia, Piscataway, NJ) were obtained from the designated suppliers. Other chemicals were purchased from Sigma (St. Louis, MO).

(2) Cell culture

KRNC cells were grown in Dulbecco's modified Eagles medium containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO<sub>2</sub>/95% air. Cells were fed every other day and passaged at 70-90% confluency with 0.1% trypsin in CA<sup>2+</sup>-Mg<sup>2+</sup>-free saline containing 0.02% EDTA. IM9 and Jurkat cells was grown in RPMI 1640 containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO<sub>2</sub>/95%. Cell viability was more than 90% as assessed by the exclusion of trypan blue.

(3) **Primer design**

Rat clones of Gα proteins (G<sub>i-1</sub> (RATBPGTPB), G<sub>i-2</sub> (RATBPGTPA), G<sub>i-3</sub> (RATBPGTP), G<sub>s</sub> (RATBPGTPD), and G<sub>0</sub> (RATBPGTPC) were retrieved from GenBank release 65.0 (HIBIIO, Hitachi America, Brisbane, CA). The nucleotide sequence similarity among these clones were then analyzed by the multiple alignment program (DNASIS, Hitachi). We have initially identified 7 highly conserved areas among them. These conserved nucleotide sequences were then analyzed against all mammalian sequences in GenBank in order to identify other similar sequences. The designed oligonucleotides (Seq. ID No. \*) were synthesized by Genosys Biotechnologies (Woodlands, TX), and suspended in water at 100 pg/ml.

(4) **PCR**

One µl of the template DNA was mixed with 1 mM each of dATP, dGTP, dCTP and dTTP, 1 µl of each PCR primers, 1 µl of 25 mM MgCl<sub>2</sub>, 5 µl PCR buffer, and 0.5 µl of Taq polymerase (18). PCR was then carried out in a DNA thermal cycler (model 480, Perkin-Elmer Cetus, Norwalk, CT) with 30 cycles of annealing temperature at ranging from 37°C to 65°C for 1.5 min, 72°C extension for 4 min followed by 95°C denaturization for 1.5 min. In separate experiments, 35% of dTTP was replaced with biotin-dUTP in order to prepare biotin-labeled probes.

(5) **Southern blot**

PCR products were separated by electrophoresis in 1.2% agarose, and stained with ethidium bromide (19). Gels were then depurinated in 0.25 N HCl for 30 minutes and denatured in 0.5 N NaOH containing 1.5 M NaCl for 30 minutes. The gels were then neutralized with 1.0 M Tris, pH 7.6 containing 1.5 M NaCl for 30 minutes. Gels were then placed onto nylon membranes (MagnaGraph, MSI, Westboro, MA) prewetted in 10X SSPE for 10 min, and DNA was transferred onto membranes by positive pressure at 75 mmHg for 60 minutes (Posiblote, Stratagene, La Jolla, CA). The DNA from the gel was then cross-linked to the membranes with ultraviolet light at 120 mJoules (Stratalinker, Stratagene), and

the membranes were incubated with hybridization buffer (ECL) containing 5% blocking reagent (ECL) and 0.5 M NaCl at 40°C for more than 1 hour. Heat denatured biotin-labeled PCR probes were then added, and hybridization was continued overnight. The membranes were washed four times for 15 minutes each time with primary wash buffer (0.5x SSPE, 36 w/v% urea, 0.4w/v% SDS) at 45-65°C, then washed twice for 5 minutes with secondary wash buffer (2x SSPE) at room temperature, and were incubated with the blocking buffer (Genius) for at least 3 hours at room temperature. Alkaline phosphatase-conjugated streptavidine (1:5,000 dilution) was then added, and incubation was continued for an additional 30-60 minutes at room temperature. The membranes were washed four times for 15 minutes with buffer A (100 mM Tris, pH 7.5, 150 mM NaCl) at room temperature, were washed for 2 minutes once with buffer C (100 mM Tris, pH 9.5, 100 mM NaCl, 50 mM MgCl<sub>2</sub>), and soaked in Lumi-Phos 530 for approximately 1-2 minutes. The membranes were then wrapped with transparency films, and chemiluminescent signals were allowed to expose X-ray films (XAR-5, Kodak, Rochester, NY) for between 10 minutes and 1 hour.

(6) mRNA preparation and cDNA synthesis

The cells were washed with phosphate buffered saline three times, homogenized in lysis buffer (FastTrack), and then incubated at 45°C for 1 hour to eliminate any RNase activity. NaCl concentrations were adjusted at 0.5 M, and an oligo (dT) cellulose tablet was added to lysis buffer. Incubation was then continued at room temperature for an additional 40 minutes. After oligo (dT) cellulose was washed with binding buffer (FastTrack) four times, bound mRNA was eluted with DEPC-treated water. Concentrations of mRNA were determined in a spectrophotometer (Hitachi, U-2000, Irvine, CA) at OD<sub>260</sub>. The first strand cDNA was synthesized from a template mRNA in the presence of 50 mM Tris, pH 8.3, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM DTT, 0.5 mM each of dATP, dCTP, dGTP, and dTTP, poly (dT) as a primer, and reverse transcriptase (Superscript) at 37°C for 1 hour. Second strand cDNA was then synthesized in the same tube, containing 25 mM Tris, pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.15 mM β-NAD<sup>+</sup>, 250 μM each of dATP, dGTP, dCTP and dTTP, 1.2 mM DTT, 65 U/ml DNA ligase, 250 U/ml DNA polymerase, and 13 U/ml RNase H (Superscript) for 2 hours at 16°C. Synthesized cDNAs were then extracted once with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), precipitated with ethanol, and resuspended in H<sub>2</sub>O.

(7) Graphic presentation

Data on Polaroid films and X-ray films was scanned by Stratascan (Stratagene) with optimization of signal-to-noise ratio, then edited with desk top publishing software (PageMaker, Aldus, Seattle, WA).

#### **EXAMPLE 6**

##### **Characterization of Cellular Components of G Protein $\alpha$ Subunits Using PCR**

###### **(1) Introduction**

Using the methods of Example 5, we have now identified and characterized two highly conserved oligonucleotide sequences among five different  $\alpha$  subunits of G proteins, SEQ ID NO:7 and SEQ ID NO:8, which can be used as PCR primers. We have found that the sequences of all the subclasses of G proteins can be amplified under the same PCR conditions using these two oligonucleotides as PCR primers, including both G protein sequences obtained from a mixture of rat G $\alpha$  protein clones and cDNAs derived from various human tissues.

Although G proteins have been analyzed extensively from a biochemical and immunological point of view using various antibodies, antibody production without any cross-reactivity among various subclasses or with high species specificity has been quite difficult to obtain. Therefore, recent experiments have focused on Northern blot analyses to identify G protein-specific mRNA from various tissues or cells in different species. However, Northern blots require experienced handling and protection from RNase contamination, in addition to a large amount of starting cellular materials.

In contrast to these conventional methods, PCR technology is more convenient and practically useful, because it requires less material than a Northern blot analysis and has great sensitivity. However, it is difficult for PCR to quantify the amount of DNA or mRNA in starting materials. Interestingly, the final PCR products obtained using the novel G-protein PCR primers of the present invention reflect the relative composition of each of the subclasses of G $\alpha$  proteins present in the starting materials. This is probably because the five different G $\alpha$  proteins cDNAs are amplified at a similar rate with a single set of PCR primers under the same PCR conditions. If known mixtures of each of the subclasses of G $\alpha$  protein clones are assayed together with unknown test samples, as shown in Figure 13, the relative composition of G $\alpha$  proteins can be determined fairly precisely. Therefore, the present method is ideal for the characterization of G $\alpha$  proteins in various tissues and cells.

Performing PCR with the primers of the present invention is also useful in clinical and diagnostic assays in the detection of disease. Since G protein abnormalities have been

associated with hereditary diseases, cancer, forms of diabetes, and other diseases, the present PCR primers for detecting and quantifying G proteins can be used to detect these diseases and assess their severity.

In the present study, we have identified  $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_s$  and  $G_0$  using the PCR primers of the present invention. Although recent cloning has identified more subclasses of G proteins, all of these newly identified G protein cDNAs showed high degree of homology to other known G proteins. Therefore, it is expected that the primers of the present invention will amplify these subclasses as well, and that the present PCR the present PCR technique can also be applied to these new G proteins. Moreover, this PCR method can be utilized to clone unique G protein genes as well.

(2) **Primers**

We have designed two 22-mer oligonucleotides ( $G_2$  and  $G_4$ --SEQ ID NOS:700 and 701) as PCR primers. As shown in Table 4 below, these oligonucleotides contain sequences which are highly conserved among five different  $G\alpha$  protein cDNAs, having only 0-4 base mismatches per sequence. No mismatch was found in the 4 bases at the 3' end of the  $G_2$ -sense and  $G_4$ -antisense sequences. Furthermore,  $G_2$  and  $G_4$  have no self-complementary sequences more than 3 base pairs in a row (data not shown). In order to analyze whether  $G_2$  and  $G_4$  are common to all the  $G\alpha$  proteins, but not to other unrelated sequences, a homology search (DNASIS) of  $G_2$  and  $G_4$  sequences was carried out against all mammalian sequences in GenBank. As a result,  $G_2$  and  $G_4$  were found to be common to all the types of  $G\alpha$  proteins and rhodopsins of various species, but less homologous to other unrelated sequences (data not shown).

TABLE 4

Two consensus oligonucleotides (G2 and G4)  
among five different cDNAs of G protein  $\alpha$  subunits

5	Consensus sequence (# of mismatch) [SEQ ID NO:]		
	G2 AGCACATTGTGAAGCAGATGA [700]	Length (bp)	G4 TGTTTGATGTGGGAGGCCAGAG [701]
10	Gi-1 AGCACATTGTGAAGCAGATGA (1) [702]	476	TGTTTGATGTGGGAGGCCAGAG (1) [805]
	Gi-2 AGCACATTGTGAAGCAGATGA (2) [703]	479	TGTTTGATGTGGGAGGCCAGAG (3) [806]
	Gi-3 AGCACATTGTGAAGCAGATGA (3) [704]	476	TGTTTGATGTGGGAGGCCAGAG (3) [707]
	Gs AGCACATTGTGAAGCAGATGA (0) [705]	524	TGTTTGATGTGGGAGGCCAGAG (3) [708]
15	G0 AGCACATTGTGAAGCAGATGA (0) [706]	479	TGTTTGATGTGGGAGGCCAGAG (4) [709]

TABLE 5

Nucleotide sequence similarity of PCR products  
among 5 different G $\alpha$  protein clones.

20	Gi-1	Gi-2	Gi-3	Gs	G0
	Gi-1	73.4	76.8	50.2	64.0
	Gi-2	73.4	67.8	52.3	65.9
	Gi-3	76.8	67.8	47.6	62.6
25	Gs	50.2	47.6	-	55.6
	G0	64.0	62.6	55.6	-

(3) PCR of rat cloned G $\alpha$  protein cDNAs

The cloned rat G<sub>i-1</sub>, G<sub>i-2</sub>, G<sub>i-3</sub>, G<sub>s</sub> and G<sub>0</sub> cDNAs were diluted at 10 ng/ $\mu$ l, and 1  $\mu$ l of DNA was mixed with 1 mM each of dATP, dGTP, dCTP and dTTP, 1  $\mu$ l of each PCR primers, 1  $\mu$ l of 25 mM MgCl<sub>2</sub>, 5  $\mu$ l PCR buffer, and 0.5  $\mu$ l of Taq polymerase as described in the Materials and Methods. PCR was then carried out with 30 cycles of annealing temperature at 45°C for 1.5 min, 72° extension for 4 min followed by 95°C denaturation for 1.5 min, 72°C extension for 4 min followed by 95°C denaturation for 1.5 min. PCR products were then separated in 1.2% agarose gel, and stained with ethidium bromide (upper panel). Marker indicates a 0.6 Kb fragment of HindIII digested  $\lambda$  DNA. DNAs were then transferred onto nylon membranes, hybridized with subclass specific biotin-labeled probes as described in the Materials and Methods. The membranes were washed with

primary wash buffer (0.5x SSPE, 36 w/v% urea, 0.4w/v% SDS) at 65°C, incubated with the blocking buffer for at least 3 hours at room temperature, then alkaline phosphatase-conjugated streptavidine (1:5,000 dilution) was added, and incubation was continued for an additional 30 min at room temperature. After extensive washing, membranes were soaked in Lumi-Phos 530 for approximately 1-2 min, then chemiluminescent signals were exposed to X-ray film for 10 min to 1 hr (lower 5 panels).

PCR was first carried out at different annealing temperatures ranging from 37°C to 65°C using the  $\lambda$  gt10 library of human HL-60 cells. As a result, PCR products were seen only at 45°C and 55°C with a size of approximately 500 bp (data not shown), which was similar to theoretical values (576 to 524 bp) (see Table 4 above). Therefore, all the PCR was then carried out at an annealing temperature of 45°C.

As shown in Figure 14, cloned rat G $\alpha$  protein cDNAs (G<sub>i-1</sub>, G<sub>i-2</sub>, G<sub>i-3</sub>, G<sub>s</sub>, G<sub>0</sub>) were successfully amplified using the same set of PCR primers (G<sub>2</sub> and G<sub>4</sub>) with a size of approximately 500 bp in 1.2% agarose gels stained with ethidium bromide. According to the computer analysis (DNASIS), the nucleotide sequences of the amplified PCR products were less homologous among five G $\alpha$  proteins with the percentage of similarity ranging from 76.8% to 47.6%. This indicates that after PCR amplification, each of the components of G $\alpha$  proteins can be identified by Southern blot analysis, even though the sizes of the PCR products generated are very similar among the five G $\alpha$  proteins. Therefore, another PCR was carried out in which 35% of the dTTP was replaced with biotin-conjugated dUTP in order to prepare subclass-specific, biotin-labeled probes. Southern membranes were then probed with these biotin-PCR products. As shown in Figure 14, these biotin-PCR probes were highly specific to each G $\alpha$  protein subclass with washing temperature at 65°C. At low stringent washing, these probes cross-hybridized with other subclasses of G $\alpha$  proteins (data not shown).

#### (4) PCR of a mixture of rat cloned G $\alpha$ protein cDNAs

G<sub>i-1</sub>, G<sub>i-2</sub>, G<sub>i-3</sub>, G<sub>s</sub> and G<sub>0</sub> cDNA were diluted at 10 ng/ $\mu$ l ( + + + ), 100 pg/ $\mu$ l ( + + ) and 1 pg/ $\mu$ l ( + ), and various combinations of these diluted cDNA were used as a PCR templates (upper panel). Then, PCR and Southern blot were carried out as described in part (3) of this Example.

By using the G<sub>2</sub> and G<sub>4</sub> sequences, all the subclasses of G $\alpha$  protein cDNA were amplified with PCR when an equal amount of G<sub>i-1</sub>, G<sub>i-2</sub>, G<sub>i-3</sub> and G<sub>0</sub> were present in test samples Figure 13, lane 4, 5, 10). However, if all the concentrations of G $\alpha$  protein cDNA are abundant, G<sub>0</sub> is less amplified (Figure 13, lane 10), probably because the number of



mismatches between  $G_0$  and  $G_4$  is higher than others between  $G_4$  and the  $G_2$  sequences. If 1 or 2 of the 5  $G\alpha$  protein cDNAs were present in smaller quantities than the others, the amounts of amplified cDNA were relatively correlated with the starting concentrations of cDNAs (Figure 13, lane 1, 2, 3, 8, 9). Furthermore, if 1 of 5 of the  $G\alpha$  proteins' cDNA is more abundant than that of the others, this G protein gene was amplified more than others (Figure 13, lane 6, 7).

(5) PCR of rat cDNAs from various tissues

Rat  $\lambda$ ZAP cDNA libraries of pituitary (P) and intestine (I), and cDNA from the mRNA of rat kidney cells (KNRK) were used as PCR templates. Then, PCR and Southern blot were carried out as described in part (3) of this Example.

Using this PCR method,  $G\alpha$  protein genes were amplified not only from cloned cDNAs, but also from various rat cDNAs (Figure 15). In  $\lambda$ ZAP cDNA libraries from rat pituitary glands and cDNA from rat kidney KNRK cells,  $G_0$  was more abundant than  $G_s$ ,  $G_{i-2}$  and  $G_{i-3}$ , and  $G_{i-1}$  was undetectable (Figure 15, lane 1, 2).  $\lambda$ ZAP cDNA library of rat intestine contained more  $G_{i-2}$ ,  $G_{i-3}$ , and  $G_s$  and  $G_0$  (Figure 15, lane 3).

(6) PCR of human cDNAs

cDNAs from the mRNA of human IM9 B-lymphocytes and Jurkat T-lymphocytes were used as PCR templates. Then, PCR was carried out with annealing temperature at 45°C (1,3) or 55°C (2, 4). Southern blot were then performed as described in part (3) of this Example.

According to sequence analyses, the PCR products of rat  $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_s$  and  $G_0$  sequence amplification exhibited a high degree of homology to human G protein cDNAs. (see Table 6 below). Furthermore, as shown in Figure 16, PCR with a pair of  $G_2$  and  $G_4$  primers could amplify 500 bp DNA from cDNAs of human IM9 and Jurkat cells. Unlike rat cDNAs (Figure 15), both IM9 and Jurkat cells contained all the subclasses of  $G\alpha$  proteins (Figure 16). However,  $G_{i-3}$  is relatively more abundant in IM9 cells, while  $G_s$  and  $G_0$  were more in Jurkat cells than IM9 cells (Figure 16).

**TABLE 6**  
Nucleotide sequence similarity of PCR products between  
rat and human G proteins

5	Length (bp)		No. of mismatch	% Similarity	
	Rat	Human			
10	Gi-1	476	476	62	87.0%
	Gi-2	479	479	44	90.8%
	Gi-3	476	476	40	91.6%
	Gs	524	482	52	89.2%
	Go	479	479	39	91.9%

It will be evident to one having ordinary skill in the art that a variety of sequences could serve as sense or antisense primers for PCR methods or as probes for the detection of mRNA as described herein. A method for identification of such sequences that are either common to a variety of G proteins or specific to a particular species is provided hereinbelow. In the preferred embodiment of this method of identification, a computer program is used to identify the sequences. Through use of such a program, we have identified a large number of both common and specific primers and probes. Provided as Tables 7 through 13 are various sense sequences identified through the use of such a program that are useful as G protein probes and primers.

All of the sequences listed in these tables are useful within the context of the PCR methods of the present invention. The complementary antisense sequences are also useful in certain aspects of the invention. As will be known having ordinary skill in the art, for common probes that are similar, but not identical to target sequences, stringency conditions can be varied (e.g. by changes in temperature and salinity) so that such probes will hybridize or fail to hybridize with a particular target sequence. Thus, also included within the present invention are sequences that are capable of hybridizing with the same sequences as either the sense sequences listed or their anti-sense counterparts.

Additional probes for G protein also include the following:

Common G protein probes

5'-CTCTGGCCTCCCACATCAAACA-3' (SEQ ID NO: 139)

5'-TCATCTGCTTCACAATGGTGCT-3' (SEQ ID NO: 140)

Specific probes (Human & Rat common)

-43-

Gi-1 5'-GTTTTCACTCTAGTTCTGAGAACATC-3' (SEQ ID NO: 141)  
Gi-2 5'-CAAAGTCGATCTGCAGGTTGC-3' (SEQ ID NO: 142)  
5'-ATGGTCAGCCCAGAGCCTCCGG-3' (SEQ ID NO: 143)  
Gi-3 5'-GTCTTCACTCTCGTCCGAAGA-3' (SEQ ID NO: 144)  
5 5'-GCCTTGGCATGCTCATAGAATT-3' (SEQ ID NO: 145)  
Gs 5'-TTCATCCTCCCACAGAGCCTTG-3' (SEQ ID NO: 146)  
Go 5'-CGCATCATGGCAGAAAGCAG-3' (SEQ ID NO: 147)

SUBSTITUTE SHEET

Table 7. G protein common primer (G2-S)

				G2-S
				AGCACCATTGTGAAGCAGATGA
5	Human			
	Gi-1	HUMGNBPAI	SEQ ID NO: 148	::T::A::::::::::::::::::::
	Gi-2	HUMGIAA	149	::::::::::C::C::::::::::::
	Gi-3	HUMGIAB	150	::::::::::::::A::::::::::
	Gs	HUMGNPAS	151	::::::::::::::::::::::::::::
10	Go	HUMGOAQ01	152	::::::::::::::::::::::::::::
	Rat			
	Gi-1	RATBPGTPB	153	::::::A::::::::::::::::::::
	Gi-2	RATBPGTPA	154	::::::::::C::C::::::::::::
	Gi-3	RATBPGTP	155	::T::T::::::::::::A::::::::
15	Gs	RATBPGTPD	156	::::::::::::::::::::::::::::
	Go	RATBPGTPC	157	::::::::::::::::::::::::::::
	Gx	RATGXA	158	::::::::::C::C::::::::::::
	Highest matched sequences in GenBank			
	HUMADECYC adenyl cyclase 159			::::::::::::::::::::::::::::
25	RATACOAlacyl-coA oxidase 160			GC::::::::G::::::::A::::::
	HUMTGASEtransglutaminase 161			CT::::::::::::::::::::CC:AC

Table 8. G protein common primer (G4-AS)

				G4-AS
SEQUENCE ID NO:				TGTTTGATGTGGGAGGCCAGAG
5	Human			
	Gi-1	HUMGNBPA	162	.....T:.....
	Gi-2	HUMGIAA	163	.....T:T:C:
	Gi-3	HUMGIAB	164	.....A:T:..A:
10	Gs	HUMGNPAS	165	.....C:..T:..C:
	Go	HUMGOAQ01	166	.....C:C:..C:
	Rat			
	Gi-1	RATBPGTPB	167	.....C:.....
15	Gi-2	RATBPGTPA	168	.....T:T:C:
	Gi-3	RATBPGTP	169	.....A:T:..A:
	Gs	RATBPGTPD	170	.....C:..C:..C:
	Go	RATBPGTPC	171	.....C:T:G:..C:
20	Gx	RATGXA	172	::G:G:.....G:..:
	Highest matched sequences in GenBank			
	HUMLDLRRL	LDL-receptor	173	CTGA:.....TACT
	MUSHEPGFA	hepatocyte growth	174	GAG:G:.....T:A:..:
25	HUMKEREP	epidermal keratin	175	AG:.....AT:..GG:..:

Table 9. Gi-1 protein specific, human-rat common primer (Gi-1)

5	SEQUENCE ID NO:			Gi-1
				GATGTTCTCAGAACTAGAGTGAAAAC
10	Human			
	Gi-1	HUMGNBPAI	176	::::::::::::::::::::::::::::::
	Gi-2	HUMGIAA	177	TT::::::::::::CT:CCCCTGTCCCCT
	Gi-3	HUMGIAB	178	::::::::::TC:G::G:::::::::G::
	Gs	HUMGNPAS	179	AC::G:C:::T:::TCCTG::C::G
	Go	HUMGOAQ01	180	::CA:C:::C:::C::G::C:::
15	Rat			
	Gi-1	RATBPGTPB	181	::::::::::::::::::::::::::::::
	Gi-2	RATBPGTPA	182	:::::G::GC:G::CC:T:::G::
	Gi-3	RATBPGTP	183	::::::::::TC:G::G:::::::::G::
	Gs	RATBPGTPD	184	A:GCACAATTA:TTA::::::::CG
	Go	RATBPGTPC	185	::CA:C:::C:::C::G::C:::
20	Gx	RATGXA	186	TCA:::GAG::C:::A:CC:::CA

Table 9A. Human-rodent common Gi-1 specific probes.

	Name in GenBank	Gi1-735 (5'-3')	Gi1-1131 (5'-3')	SEQ ID NOS: (respectively)
10	Rat Gs RATBPPIP Gi-1 RATBPPIP Gi-2 RATBPPIP Gi-3 RATBPPIP Go RATBPPIP	A-GCACAATTA-TTA -----CG -----G-CC-G-CC-T -----TC-G-G --CA-C---C---G-C	AACAGAA-A-----AA-A---G---T -----G-----G-----C A-CA-----G-----T---G---CCG ACA-----AG-AAA---CCGCTC	224, 225 226, 227 228, 229 230, 231 232, 233
15	Human Gi-1 HUMGNBPA GenBank** HUMCPH192 MUSGBPA	-----G -----C---C---G---G---A-----G	-----A-C---CC-T---C-----	234, 235 236 237

\* : Size of PCR products using Gi1-735 and Gi1-1131 as primers.

\*\* : The highest matched sequences in GenBank (release 68.0)  
next to the G protein DNAs.

Table 10. Gi-2 protein specific, human-rat common primer (Gi-2)

5		SEQUENCE ID NO:		Gi-2 GCAACCTGCAGATCGACTTTG
	Human			
	Gi-1	HUMGNBPAI	187	:G:GGT::A::::A:::::
	Gi-2	HUMGIAA	188	: :: :: :: :: :: :: :: :: ::
10	Gi-3	HUMGIAB	189	:ACGG::AA::::T:::::
	Gs	HUMGNPAS	190	AGT:::A::T:::T:::G::
	Go	HUMGOAQ01	191	CTTCT: :: :: :: G: TG: :: :: C
	Rat			
	Gi-1	RATBPGTPB	192	:G:GAT::A:A:::: :: :: ::
15	Gi-2	RATBPGTPA	193	: :: :: :: :: :: :: :: :: ::
	Gi-3	RATBPGTP	194	AA:::G:::::T:T:TT:::
	Gs	RATBPGTPD	195	AGT:::A::T:::: :: :: G::
	Go	RATBPGTPC	196	ATTCT: :: :: :: A: TG: :: :: C
20	Gx	RATGXA	197	:T:::A:C:::T:T:T:::C:



Table 10A. Human-rodent common Gi-2 specific probes.

	Name in GenBank	Gi2-742 (5'-3')	Gi2-1102 (5'-3')	SEQ ID NOS: (respectively)
Rat				
GS	RATBPCTPD	G-C-A-A-.....ATCAAG-	---C--G-----CG-GC-ACGC	238, 239
Gi-1	RATBPCTPB	--A-----C-----A--C-	---A--T-----C-----	240, 241
Gi-2	RATBPCTPA	.....	.....	242, 243
Gi-3	RATBPCTP	--A-----A-----C-	---G-----T-----T-	244, 245
Go	RATBPCTPC	-G-C-----C--T--G--C-	CTA-----G--C-CTCA-C	246, 247
Mouse				
Gi-2	MUSGI	.....	.....	248, 249
Human				
Gi-2	HUMGIR	.....	.....	250, 251
GenBank**				
	HUMCHPF	.....C-----TGG	.....A-----A-A-G---T	252
	MUSHSP84A			253

\* : Size of PCR products using Gi2-742 and Gi2-1102 as primers.  
 \*\* : The highest matched sequences in GenBank (release 68.0)  
 next to the G protein DNAs.

Table 11. Gi-3 protein specific, human-rat common primer (Gi-3)

5				Gi-3
		SEQUENCE ID NO:		TCTTCGGACGAGAGTGAAGAC
	Human			
	Gi-1	HUMGNBPAI	198	:::CA:A::T::::::::::A::
10	Gi-2	HUMGIAA	199	G::A:::::CC:C::A:::::
	Gi-3	HUMGIAB	200	::::::::::
	Gs	HUMGNPAS	201	GGGAAATCGA:::T:::G:::
	Go	HUMGOAQ01	202	:::C:::::GA:::CGTG:A::
15	Rat			
	Gi-1	RATBPGTPB	203	:::CA:A::T::::::::::A::
	Gi-2	RATBPGTPA	204	G::G:::::CC:T::::::::::
	Gi-3	RATBPGTP	205	::::::::::
	Gs	RATBPGTPD	206	TTCCT:::A:::T:::T:TG
20	Go	RATBPGTPC	207	CGCAT:::G::C:C::::CCA
	Gx	RATGXA	208	AAC::::G:A::::CACCAT:

Table 11A. Human-rodent common Gi-3 specific probes.

	Name in GenBank	Gi3-407 (5'-3')	Gi3-730 (5'-3')	SEQ ID NOS: (respectively)
5		TTGTTTGTAGCTGGCAGTCTGA	GAGGAGTGACAGCAATTATCT	
10	Rat			
	Gs	RATBPGIPD	CC-CAAGT-GATC-----TC	254, 255
	Gi-1	RATBPGIPB	A-----GA-----GGC-CTA-T	256, 257
	Gi-2	RATBPGIPA	CCTACAC---A-AT-----C	258, 259
15	Gi-3	RATBPGIP	-----T---G---C---C---	260, 261
	Go	RATBPGTPC	CG-----T---AGTC-TTAC-	262, 263
	Human		TC-----AT-----TCAACGAC	
	Gi-3	HUMGTPBP	-----	264, 265
	GenBank**			
20	HUMAP08	G-----C---A-----T-----C		266
	HUMINSR02		C-----G---AC---T---	267

\* : Size of PCR products using Gi3-407 and Gi3-730 as primers.  
 \*\*: The highest matched sequences in GenBank (release 68.0)  
 next to the G protein DNAs.

Table 12. Gs Protein specific, human-rat common primer (Gs)

5				Gs
	SEQUENCE ID NO:			AATTCTATGAGCATGCCAAGGC
10	Human			
	Gi-1	HUMGNBPAI	209	GGGCGG::::T:::C::::CT
	Gi-2	HUMGIAA	210	:::ATG:::::GCA::::GCTA
	Gi-3	HUMGIAB	211	C:G::::ACTA:::T::::CTC
	Gs	HUMGNPAS	212	::::::::::

15

Table 12A. Human-rodent common Gs specific probes.

	Name in GenBank	Gs-246 (5'-3')	Gs-824 (5'-3')	SEQ ID NOS: (respectively)
5				
10	Rat			
	Gs			
	Gi-1	RATBPCTPD	-----	268, 269
	Gi-2	RATBPCTPB	CG--G--G-T-----CCGCA	270, 271
	Gi-3	RATBPCTPA	CAA-GG-----CA-A-----	272, 273
	Go	RATBPCTP	TCAGGA-----T-AA-----T	274, 275
15	Mouse	RATBPCTPC	TTTGG-G-G-----TA-----T	276, 277
	Gs-1	MUSGS	-----	278, 279
	Gs-2	MUSGTPAMU	-----	280, 281
	Human			
	Gs-1	HUMGSA1R	-----	282, 283
20	Gs-2	HUMGSA2R	-----	284, 285
	GenBank**			
	HUMBP2	A-----A-C-----A-----	A-----T-----T-----A-----	286
25	RATAFPM			287

\* : Size of PCR products using Gs-246 and Gs-824 as primers.

\*\* : The highest matched sequences in GenBank (release 68.0)  
next to the Gs protein DNAs.

Table 13. Go Protein specific, human-rat common primer (Go)

5		SEQUENCE ID NO:		Go CTGCTTTCTGCCATGATGCG
	Human			
	Gi-1	HUMGNBPAI	213	GAC:::::TAA:::TGACA
	Gi-2	HUMGIAA	214	:::TGA:::C:A:G:C:::
10	Gi-3	HUMGIAB	215	:::A:::AGA:CTTCACA
	Gs	HUMGNPAS	216	TCAACGA:::::G:::CATC
	Go	HUMGOAQ01	217	:::~::~~::~~::~~::~~::~~::
	Rat			
15	Gi-1	RATBPGTPB	218	:::A:::CCA:G:T:::TT
	Gi-2	RATBPGTPA	219	T:::GA:CC:::GCG:::
	Gi-3	RATBPGTP	220	:::~::~~::~~::~~::~~::~~::
	Gs	RATBPGTPD	221	:::~::~~::~~::~~::~~::~~::
	Go	RATBPGTPC	222	:::~::~~::~~::~~::~~::~~::
20	Gx	RATGXA	223	:A:::C:T:::C:::CTG:

Table 13A. Human-rodent common Go specific probes.

	Name in GenBank	Go-1224 (5'-3')	Go-1397 (5'-3')	SEQ ID NOS. (respectively)
5		AGGACATCTCCGAACGAG	TATGACCAGGTCTCCACG	
	Rat			
	Gs	RATBPGTDP		
	Gi-1	RATBPGTPB		288, 289
	Gi-2	RATBPGTPA		290, 291
	Gi-3	RATBPGTP		292, 293
	Go	RATBPGTPC		294, 295
	Mouse			296, 297
	Go-A	MUSGOASA		
	Go-B	MUSGOASB		298, 299
10	Hamster			300, 301
	Go-1	HAMHITAO1		
	Go-2	HAMHITAO2		302, 303
	Human			304, 305
15	Go	HUMGOA1(2)		
	GenBank**			306, 307
20	MUSPC3AA	G-----T--TC-----		308
	RATACACA		-----A-GT	309
25				

\* : Size of PCR products using Go-1224 and Go-1397 as primers.  
 \*\*\*: The highest matched sequences in GenBank (release 68.0)  
 next to the G protein DNAs.

**Substance P**

Another example of clinically useful measurement of mRNA would be for the mRNA of substance P. Substance P is a neurotransmitter. Because the nerves expressing substance P are involved in pain receptor pathways, the ability to measure the mRNA of substance P would be useful in the development of analgesics.

**EXAMPLE 6****Detection of Human Substance P Receptor mRNA From Substance P Receptor cDNA-Transferred Cells, Vector-Transfected Cells, and Evaluation of Their Reliability****(1) Preparation of the First Nucleotide Probe**

The first nucleotide probe used for quantification of substance P receptor-specific mRNA was an oligodeoxyribonucleotides containing a 5' amino group. This probe was manufactured using a DNA synthesizer (model 380B, Applied Biosystems, CA) and is shown below:

5'-NH<sub>2</sub>-GGA CTT ATG AGA AAG CGT ACC A-3' (Seq ID NO:5).

**(2) Immobilization of the Second Nucleotide Probe Onto Insoluble Supports**

The first nucleotide probe were dissolved in DEPC-treated water at a concentration of 1 µg/µl, then mixed with the EDC/Sulfo-NHS solution described in Example 1 Paragraph (2) in a ratio of 1:25 (Vol:Vol). 50 µl of the above mixed solution was added to each well of a Sumilon microtiter plate (MS-3796F, Sumitomo Bakelite, JAPAN) and the plates were incubated overnight at room temperature. Following incubation, the reaction solution was removed with an aspirator.

**(3) Preparation of Standard mRNA**

A plasmid vector pKR2 containing human substance P receptor cDNA (provided by Dr. S. Nakanishi, Kyoto University, Japan) was linearized by SmaI (Promega, WI) digestion. Substance P receptor-specific mRNA was prepared by the method described in Example 1 Paragraph (3) using T7 RNA polymerase (Promega, WI). Concentrations of the resultant mRNA were determined by spectrophotometry and then diluted with DEPC-treated water to prepare standard mRNA solutions as described in Example 1 Paragraph (3).

**(4) Preparation of Cell Lysates**

One line of human T lymphocyte-derived Jurkat cells which had been transfected with substance P receptor cDNA was examined for transcription of substance P receptor mRNA (hereafter called JSP, provided by Dr. D. G. Payan, University of California, San Francisco, CA). Another line of Jurkat cells which had been transfected with a vector that



didn't express substance P receptor (hereafter called JVec, a kind gift of Dr. D.G. Payan, ibid) was used as a negative control. Cells were collected by centrifugation, washed with PBS, and resuspended in lysis buffer at  $1 \times 10^7$  cells/ml as described in Example 2. The cells were sheared by repeated passage through a 21 gauge needle, then incubated in a slow-shaking waterbath at 45°C for 1 hour.

(5) **Hybridization of the First Nucleotide Probe to mRNA**

In order to remove possible RNase contamination in the microtiter wells containing the bound first nucleotide probe, 250  $\mu$ l of lysis buffer was added and incubated at 45°C for 1 hour as described in Example 1. The buffer aspirated off, and 50  $\mu$ l of lysis buffer containing 0.5 M NaCl and various concentrations of standard human substance P receptor-specific mRNA was added into each well, and incubated at 50°C for 1 hour to allow hybridization as described in Example 1.

(6) **Hybridization of the Labeled Second Nucleotide Probe to a Poly(A) Tail of mRNA, and Measurement of Chemical Activities of the Labeled Second Nucleotide Probe**

After hybridization of the first nucleotide probe to bound mRNA as described in Paragraph (5) above, the hybridization solution was removed by aspiration and individual wells were washed once with 250  $\mu$ l of lysis buffer. Five  $\mu$ l of lysis buffer containing 0.5 M NaCl and 1  $\mu$ l of the biotinylated second nucleotide probe solution (35 pmol/ $\mu$ l) was added into each well and incubated at room temperature for an additional 1 hour to allow the second hybridization as described in Example 1.

Chemical activities of the labeled second nucleotide probe was also measured as described in Example 1.

(7) **Preparation of Calibration Line**

The calibration line of standard substance P receptor-specific mRNA is shown in Fig. 9.

(8) **Evaluation of Reproducibility**

Concentrations of substance P receptor-specific mRNA in JSP and JVec cell lysates were measured twice. Table 7 shows concentrations of mRNA obtained from the calibration line as described in Paragraph (7). Cell numbers from each sample were counted by a phase hemacytometer (American Scientific Products, IL). As shown in Table 14, substance P receptor-specific mRNA concentrations were found in JSP cells that had been transfected with substance P receptor cDNA, but were not detected from JVec cells transfected with only a control vector. This shows that the invented method provides good quantification of a specific mRNA with high reliability.

TABLE 14

Concentrations of Substance P Receptor-Specific mRNA in Jurkat Cells Transfected with Substance P Receptor cDNA or Only a Control Vector.

No. of Assay	mRNA Concentration	
	JSP [pg/cell] (Transfected)	JVec [fg/cell] (Control)
1	135	<0.1
2	62.4	<0.1

Oligonucleotide Primers and Probes for the Detection of Jun Oncogenes

One aspect of the present invention relates to DNA primers and probes for the detection of jun oncogenes, such as by PCR. Expression of the jun gene can serve as a marker for cell growth.

It is well known that certain oncogenes, such as jun or fos are most rapidly expressed when cells are stimulated with outside stimuli to proliferate. Therefore, detection or quantification of levels of expression of jun or fos oncogenes will be a good marker for cellular mitogenic activity.

Jun oncogenes are one of oncogenes first reported by Maki, et al. (Proc. Natl. Acad. Sci. USA, 84:2848-2852, 1987), and current intensive molecular cloning identified at least three subtypes of jun oncogenes, such as jun-B, c-jun and jun-D. Based on nucleotide sequence comparison, three subtypes of jun oncogenes in both humans and mice are different (Ryder K, Nathans D. Proc. Natl. Acad. Sci. USA 85:8464-8467, 1988; Ryder K, et al., Proc. Natl. Acad. Sci., USA 85:1487-1491, 1988; Hattori K, et al., Proc. Natl. Acad. Sci. USA 85:9148-9152, 1988; Schuette J, et al., Cell, 59:987-997, 1989). However, because it is still unclear how these three different jun oncogenes are involved in cell growth, one has to analyze all of the three genes during cell growth.

Northern blot analysis is widely accepted in order to detect specific genes (Sambrook J, et al., Molecular cloning. A laboratory manual, 2nd eds. pp. 7.39-7.52), and some investigators have already adapted this method to detect jun oncogenes (Sherman, et al., Proc. Natl. Acad. Sci. USA, 87:5663-5666, 1990; Oursler MJ et al., Proc. Natl. Acad. Sci.

USA, 88:6613-6617, 1991). In this method, mRNA is first purified from tissues/cell of researchers' interest, separated on agarose gel electrophoresis. After electrophoresis, mRNA is transferred onto membranes, and hybridized with radioactive probes to identify positive band(s) on autoradiography. However, this method is not so sensitive that it is difficult to identify specific signals if materials have only small quantity of genes. Alternatively, because reverse PCR ("Molecular Cloning") is also used to detect a wide variety of genes from different tissues/cells, this technique is also applicable for the detection of jun oncogene. In this method, mRNA is first convert to cDNA by reverse transcriptase, then specific gene fragments are amplified by PCR by using one set of primers (sense and anti-sense primers). Amplified gene can be seen on agarose gel electrophoresis in the prepense of ethidium bromide. This method is much more sensitive than the conventional Northern blot analysis, but in order to detect three different subtypes of jun oncogenes, one has to prepare six primers. Also, because of different nucleotide sequences between humans and mice, one has to prepare human and mice primers separately.

One goal we accomplished with the present invention was the design of oligonucleotide sequences for both sense and anti-sense primers that could be readily used for detecting the minuscule amounts of the jun gene found *in vivo*. We also wanted these primers to be useful for detecting both human and murine jun sequences. Thus, we selected primers which fulfilled the below conditions:

- (a) A common nucleotide sequence among three subtypes of jun oncogenes (jun-B, c-jun and jun-D), with maximum 4 base mismatches among them;
- (b) A common nucleotide sequence among humans and mice, with maximum 4 base mismatches between them;
- (c) At least 5 bases from 3' end are 100% identical without any mismatches among three subtypes and between humans and mice;
- (d) A nucleotide sequence length in both sense primer and anti-sense primer from 17 to 50;
- (e) A difference of  $T_m$  between sense primer and anti-sense primer is within 2 degree C;
- (f) Lack of complementary structure with more than 4 bases long in either the sense or the anti-sense primer;
- (g) Lack of complementary structure with more than 4 bases long between the sense and the anti-sense primer;

- (h) Both the sense and the anti-sense primer exist in the area of coding sequence;
- (i) The length of DNA to be amplified is greater than 200 bases; and
- (j) The nucleotide sequence homology of amplified genes among three subtypes of jun oncogenes is not greater than 80%, and after gene amplification by the PCR, one can characterize each subtype of jun ongenes by Southern Blot.

DNA fragments (both sense and anti-sense primers) which satisfy the conditions as shown above were investigated and the several candidate oligonucleotides were synthesized. The test results showed the sense primer 5'-CCCTGAAGGAGGAGCCGCAGAC-3' and anti-sense primer 5'-CGTGGGTCAAGACTCTGCTTGAGCTG-3' were desirable candidates.

Table 15 shows the nucleotide sequence homology between a preferred sense primer and three subtypes of jun oncogenes in both humans and mice.

TABLE 15

Sense primer			
5'-CCCTGAAGGAGGAGCCGCAGAC-3' (SEQ ID NO:310)			
Mouse	jun-B	--T-T--A-----	(SEQ ID NO:311)
	c-jun	-----A-----	(SEQ ID NO:312)
Human	jun-B	--T-C-----A-----	(SEQ ID NO:313)
	c-jun	-----T-----	(SEQ ID NO:314)

-: indicates identical base to the sense primer.

In addition, Table 16 shows the nucleotide sequence homology between the sense primers and three subtypes of jun oncogenes in both humans and mice.

TABLE 16

		Antisense primer	
		3'-GTCGAGTTCGTCTTTCAGAACTGGGTGC-5' (SEQ ID NO:315)	
5	Mouse jun-B	-----C-----A-	(SEQ ID NO:316)
	c-jun	-----T----	(SEQ ID NO:317)
	jun-D	----T-----G-G-C----	(SEQ ID NO:318)
10	Human jun-B	----T----C-----	(SEQ ID NO:319)
	c-jun	----T-----	(SEQ ID NO:320)

--: indicates identical base to the sense primer.

Also, Table 17 shows nucleotide sequence homology of amplified genes between each of subtypes of jun oncogenes.

TABLE 17

		Subtypes	Homology (%)
20	jun-B & c-jun		72.0
	jun-B & jun-D		73.6
	c-jun & jun-D		75.3

Table 18 shows the homology between human and mouse subtypes.

TABLE 18

		Subtypes	Homology between human & mouse (%)
30	jun-B		93.0
	c-jun		91.9

The primer sequences can be modified in a variety of ways. For example, a primary amine residue, nucleotide sequence recognized by a restriction enzyme, or RNA promoter sequences for further modification after PCR amplification, can be added to the 5' end of the primers. These modifications are represented herein by the designation "X". Thus, for example, a modified primer as shown in Table 16, can be represented as 5'-XCGTGGGTCAAGACTTCTGCTTGAGCTG-3' can also be used, in which X represents the primary amine residue, nucleotide sequence recognized by the restriction enzyme, or RNA promoter sequences for further modification after PCR amplification.

The attachment of a restriction site at the 5' end (X) is useful for cloning of amplified genes. The attachment of RNA promoter sequences at 5' end (X) is useful in

RNA transcription and RNA transcription-based amplification. The attachment of primary amine at the 5' end is useful for coupling reaction to labeling compounds or solid supports, so that the amplified gene can be easily quantified.

For the above-mentioned nucleotide sequence (X) which undergoes cutting by a restriction enzyme, instead of those nucleotide sequences with higher frequency such as Eco RI or Bam HI, it is often desirable to use longer recognition sequences (e.g. 8 bases) which appear in lower frequency such as the nucleotide sequence, 5'-GCGGCCGC-3', which is recognized by NotI or the nucleotide sequence, 5'-TTAATTAA-3', which is recognized by PacI. The reason for this is the relatively common nucleotide sequences recognized by EcoRI or BamHI can result in false positive reaction during PCR amplification because of priming of primers to undesirable DNA sequences.

The above-mentioned nucleotide sequence (X) containing an RNA promoter can be any of a variety of such sequences, such as the T7, SP6 or T3 RNA promoter sequences.

The above-mentioned primary amine residues (X) can be added at the 5' end during oligonucleotide synthesis.

The DNA fragments for the sense primer and anti-sense primer in this invention can be easily synthesized in DNA synthesizer which are available in the market. These synthesized oligonucleotides can be purified by high pressure liquid chromatography or gel electrophoresis.

The test material to be analyzed is usually a total RNA or purified mRNA from cells or tissues. If desired, cells or tissues can be tested in their natural state without any pretreatment; however, in the case of drug testing, this can be accomplished by reacting the drug with cells or tissues in the test tube, or the drug can be administered (intravenous injection, subcutaneous injection, intramuscular injection, oral administration, intra-abdominal injection) to a laboratory animal, and after a set time period, cells or tissues can be removed for RNA/mRNA purification.

Methods of purification of total RNA or mRNA are well known in the art. Examples of such methods are the standard protocol described in Sambrook, et al., "Molecular Cloning, A Laboratory Manual, 2nd Ed.," Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), and a commercially available kit such as FastTrack from Invitrogen (San Diego, CA). In either case, in order to avoid introducing RNase, hands should preferably be protected with vinyl gloves, and the instruments used for experiments not be touched with bare hands. Also, any glass containers to be used in the experiments should be heated prior to use at approximately 250 degrees C for at least 4 hours. Furthermore, 0.1% diethyl

pyrocarbonate (hereafter abbreviated DEPC) is added to the water which is to be used, and after incubation at 37 degrees C overnight, treated water is autoclaved. Other basic procedures are undertaken as outlined in the "Molecular Cloning," pp. 7.3-7.5. Vanadyl ribonucleoside complex can also be added to samples as described above.

5 The method of synthesis of cDNA from mRNA template using reverse transcriptase is described in the "Molecular Cloning," pp. 8.11-8.13.

For PCR amplification, synthesized cDNA is mixed with sense primer, antisense primer, 4 types of deoxynucleotides (dATP, dCTP, dGTP and dTTP), Taq polymerase, inorganic salts, and other necessary materials, and PCR reaction is undertaken in the  
10 Thermal cycler (Perkin-Elmer Cetus). A typical PCR method is described in "Molecular Cloning," pp. 14.2-14.33).

In order to analyze the amplified gene, it is appropriate to use electrophoresis. After the amplified gene undergoes electrophoresis in agarose gel, the DNA is stained with ethidium bromide. The amplified DNA band will be then visible under fluorescent light as  
15 described in "Molecular Cloning," pp. E.3-E.4. After taking photographs of the DNA band, it is also possible to quantify the intensity of each bands by scanning photographs and analysis of scanned picture by commercially available system, such as Stratascan (Stratagene, La Jolla).

Furthermore, after agarose gel electrophoresis, amplified genes can be transblotted  
20 onto membranes, and subtypes specific genes can be detected by hybridizing with labeled probes followed by exposing labeled signals, such as 32P or chemiluminescence to either Polaroid films or X-ray films (Southern blot).

#### EXAMPLE 7

##### Synthesis of Sense and Anti-Sense Primers and 25 Amplification of Mouse Clones of Jun Oncogenes

The sense primer SEQ ID NO:9 (s943-2) and the anti-sense oligonucleotide SEQ ID NO:10 (AS1132-2) were synthesized by a synthesizer 380 B type (Applied Biosystems Co.). After treatment with ammonium hydroxide at 55°C overnight, synthesized oligonucleotides were dried in a Speed-Vac (Savant Co.), and the concentration was  
30 adjusted at 1 microgram/ml with water, and was stored at -20 degrees C until use.

1 microliter of one of the three types of mouse jun clones (jun-B, c-jun, or jun-D, obtained from ATCC), 1 microliter of sense primer, 1 microliter of anti-sense primer, 5 microliters of 10X buffer for PCR (Promega), 1 microliter of 25 mM magnesium chloride, 4 microliters of 10 mM dNTP mix, and 0.5 microliters of Taq polymerase (Promega) were

-64-

missed, and water was added up to a total volume of 50 microliters. After adding of two drops of mineral oil, PCR was undertaken using the thermal cycler, model 480. After reaction mixture was heated at 95 degree C for 10 min, PCR was carried out with the following cycles 30 times: annealing at 55 degrees C for 1.5 minutes, extension at 72 degrees C for 4 minutes, and denaturing at 95 degrees C for 1.5 minutes.

After PCR, 10 microliters of the sample were mixed with 1 microliter of 10x loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, and 15% Ficoll, Type 400), and electrophoresis was carried out in 1.5% agarose gel containing 5 microgram/ml ethidium bromide. After electrophoresis, the amplified DNA bands were visualized by an ultraviolet light.

As a result, it was found that in each of the cases where the clones jun-B, c-jun, and jun-D were used, a single band for the amplified DNA was observed at the position of approximately 270 bp. This indicates that the above-mentioned set of primers can recognize and amplify all three types of mouse jun oncogenes, such as jun-B, c-jun, jun-D.

(1) Alternative Synthesis of PCR primers for amplification of jun gene specific mRNA

Sense (jun-s) and antisense (jun-as) oligodeoxyribonucleotides specific for jun oncogenes were synthesized in a DNA synthesizer as described above. The nucleotide sequences were as follows:

jun-s 5'-CCCTGAAGGAAGAGCCGCAGAC-3' (Seq ID NO: 11)

jun-as 5'-CGTGGGTCATGACTTTCTGCTTGAGCTG-3' (Seq ID No: 12)

(2) PCR amplification of cDNA

To 1  $\mu$ l of ss-cDNA as described in paragraph (IV), 0.1  $\mu$ g (1  $\mu$ l) each of jun-s and jun-as oligonucleotides, 5  $\mu$ l of 10X PCR buffer (Promega), 1  $\mu$ l of 25 mM  $MgCl_2$ , 4  $\mu$ l of 10 mM dNTP mixture (Promega), 0.5  $\mu$ l Taq polymerase (Promega), and 36.5  $\mu$ l of DEPC-water were mixed. Two drops of mineral oil were overlaid onto the reaction mixture to prevent evaporation. The reaction mixture was first heated at 95°C for 10 minutes, then PCR was carried out in a Thermal cycler (Model 480, Perkin-Elmer Cetus) with 30 cycles of annealing at 55°C for 1.5 minutes, extension at 72°C for 4 minutes and denaturing at 95°C for 1.5 minutes as described above.

Ten  $\mu$ l of the resultant PCR products were then mixed with 1  $\mu$ l of 10X loading buffer, and applied to a 1.5% agarose gel containing 5  $\mu$ g/ml ethidium bromide. Electrophoresis was carried out in a 1X TBE buffer at 100 V/6.5 cm width for approximately 1 hour. After electrophoresis, fluorescent DNA bands were recorded onto Polaroid film. Lanes 1-8 of Figure 17 indicate duplicate experiments of a 1:1000 dilution.



full strength (equivalent to 1 ml of blood), 1:10, and 1:100 dilution respectively of a leukocyte lysate. As shown in Figure 17, the jun oncogene was amplified from the 1:10 and 1:100 dilution displaying an estimated size of 370 base pairs. This size is similar to the 372bp theoretical value.

5

### **EXAMPLE 8**

#### **The effect of pretreatment with EGF on expression of jun oncogenes in human mononuclear leukocytes**

##### **(1) Pretreatment of human leukocytes with EGF**

40 ml of phosphate buffered saline (PBS) is added to 20 ml of heparinized human blood and mixed. 10 ml each of this sample is overlaid onto 3 ml of IsoLymph, then centrifuged for 30 minutes at 400 x g. After washing the pellet three times with PBS, the pellet is resuspended in 3 ml of PBS. 1 ml each of this sample is then placed in three tubes (No. 1 to No. 3). 1 ml of PBS is placed in the fourth tube as control. The total of four tubes are incubated for 10 minutes at 37 degrees C, then EGF (Epidermal Growth Factor) is placed in tube No. 1 at a final concentration of 30 ng/ml. After 15 minutes, EGF is placed in the same manner in tube No. 2, and incubated for another 5 minutes. After 5 min, RNA is extracted from all 4 tubes simultaneously. Thus, the time period of the pretreatment of human leukocytes by EGF is 20 minutes for tube No. 1, 5 minutes for tube No. 2, 0 minutes for tube No. 3.

20

##### **(2) Extraction of RNA from cells**

The above four tubes are taken out and undergo centrifugation by microfuge for 10 seconds. Then, the supernatant is discarded and the below described lysis buffer is added to the pellet, and after thorough mixing, this is incubated for 30 minutes at 45 degrees C.

Contents of lysis buffer:

25

10 mM EDTA pH 8.0

0.5% SDS (bacteria removed by non-bacterial filter)

0.2 M NaCl

DEPC treated water

RNA inhibitor 500 unites/ml

30

Vanadyl Complex 10 mM

Proteinase K 200 microgram/ml

##### **(3) Purification of mRNA**

After 5 M NaCl is added to the above cell lysates to obtain a final concentration of 0.5 M, oligo (dT) cellulose (Stratagene Co.) is added, and undergoes a reaction for 30

minutes at room temperature. Then, after washing the cellulose in 10 ml binding buffer (20 mM Tris-HCl, pH 7.6, 1 mM EDTA, 0.5 M NaCl) 5 times, 0.35 ml of DEPC treated water is added, and mRNA is eluted from the solid phase. Then, 53 microliters of 2 M sodium acetate and 2.5 times the volume of ethanol are added, and after cooling in a dry ice for 20 minutes, this undergoes centrifugation at 15,000 rpm for 20 minutes. After washing the pellet in 75% ethanol one time, this is dried, and then dissolved in 10 microliters of DEPC treated water.

(4) Synthesis of cDNA

50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM magnesium chloride, 10 mM DTT, 0.5 mM dNTP (dATP, dCTP, dGTP, dTTP), 50 micrograms/ml oligo (dT) primer, and 10,000 units/ml reverse transcriptase are added to 10 microliters of mRNA obtained from above to a total volume of 20 microliters, and this undergoes a reaction for one hour at 37 degrees C. After the reaction, 20 microliters of a phenol:chloroform:isoamyl alcohol mixture is added, and cooled for 20 minutes in a dry ice to precipitate the cDNA. After centrifugation for 20 minutes at 10,000 rpm, the pellet is washed one time in 75% ethanol. Then, after drying, this is dissolved in 20 microliters of autoclaved water, and stored at -20 degrees C.

(5) PCR

1 microliter each of sense primer and anti-sense primer (1 mg/ml) for jun gene amplification are added with 2 microliters of cDNA. After mixing this with 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 2.0 mM magnesium chloride, 100 micrograms/ml gelatin, and 0.2 mM dNTP, 2.5 units of Taq polymerase are added (final volume is 50 ml). After reaction mixture was heated at 95 degree C for 10 minutes, PCR was carried out with the following cycles 30 times: annealing at 55 degrees C for 1.5 minutes, extension at 72 degrees C for 4 minutes, and denaturing at 95 degrees C for 1.5 minutes.

(6) Agarose gel electrophoresis

After completing PCR, 10 microliters of the reacted solution is taken and undergoes electrophoresis in the same method as described in Example 7. The results showed that in the leukocytes which did not undergo treatment with EGF (i.e., sample which underwent 0 treatment time in tube No. 3) was found to have a minimal band for the amplified DNA at the position of about 270 bp size. However, it was found that the band of amplified DNA was increased after 5 min, then returned to the basal levels within 20 min.

EXAMPLE 9

Effect of pretreatment with PHA on jun gene expression in human leukocytes

In the place of EGF, PHA (at a final concentration of 10 micrograms/ml) was utilized and the pretreatment times were set at 0 minutes, 5 minutes, 15 minutes, and 30 minutes. Other procedures were identical to those of Example 8. As a result, with the pretreatment with PHA done at 15 minutes, it was found that the band for the amplified DNA at the position of approximately 270 bp was maximized, but (the time periods) afterwards, this band decreased.

As with the other mRNAs that can be analyzed in accordance with the methods of the present invention, a variety of jun sequences could serve as sense or antisense primers for PCR methods or as probes for the detection of mRNA as described herein. A method for identification of such sequences that are either common to a variety of jun mRNAs or specific to a particular species is provided hereinbelow. In the preferred embodiment of this method of identification, a computer program is used to identify the sequences. Through use of such a program, we have identified a large number of both common and specific primers and probes. Provided as Tables 19 through 24 are various sense sequences identified through the use of such a program that are useful as jun gene probes and primers.

All of the sequences listed in these tables are useful within the context of the PCR methods of the present invention. The complementary antisense sequences are also useful in certain aspects of the invention. As will be known by those having ordinary skill in the art, for common probes that are similar, but not identical to target sequences, stringency conditions can be varied (e.g. by changes in temperature and salinity) so that such probes will hybridize or fail to hybridize with a particular target sequence. Thus, also included within the present invention are sequences that are capable of hybridizing with the same sequences as either the sense sequences listed or their anti-sense counterparts. Additional probes for jun genes include the following:

Common jun gene probes

5'-CCATGTCGATGGGGGACAGCGG-3' (SEQ ID NO: 13)

5'-CTGTTTAAGCTGCGCCACCTG-3' (SEQ ID NO: 14)

5'-GTCTGCGGCTCCTCCTTCAGGG-3' (SEQ ID NO: 15)

5'-CGTGGGTCAAGACTTTCTGCTTGAGCTG-3' (SEQ ID NO: 16)

Specific probes

B type: 5'-CACTTGGTGGCCGCCAG-3' (SEQ ID NO: 17)

C type: 5'-GAGCATGTTGGCCGTGG-3' (SEQ ID NO: 18)

Human D type: 5'-GATGCGCTCCTGCGTGT-3' (SEQ ID NO: 19)

Mouse D type: 5'-GCCTGTTCTGGCTTTTGAGGG-3' (SEQ ID NO: 20)

Table 19. Jun-common sense primer (S943-2).

Locus	GenBank	Pos	Sequences (5'-3')
			CCGCTGTCCTCCCATCGACATGG
5	Human		
	B: humjunc	1189	---G-----A-----
	C: humjuna	1981	---C-----
10	D: humjundr	943	---T---G-----
	Mouse		
	B: musjunba	1079	---TG-----A-----
	C: musjunc	1344	---C-----T-----
15	C: muscjunc	1646	---C-----T-----
	C: musjun	1084	---C-----T-----
	D: musjund	927	---G-----
	D: musjunda	782	---G-----
20	D: musjundr	793	---G-----
	Rat		
	C: atjunap1	1082	---CT-----
	C: ratrig9	2984	---CT-----
25	Chicken		
	C: chkjun	1470	---C-----T-----
	Quail		
	C: quljun	1186	---C-----T-----
30	Drosophila		
	C: drojun	1038	A-CG-TAAT-----T-----
35	Highest matched sequences in EMBL		
	SDHAM2G Yeast NAM2 gene (SEQ ID NO:36)		
	PRK2TRFB Plasmid PK2 trfb ope (SEQ ID NO:37)		
	DMSYT D.melanogaster synap (SEQ ID NO: 38)		
			-----A-A-----GAAT
			GT-----GC-T-
			-----G-A-----UC-----



-71-

Table 21. Jun-B specific probe (B-1258).

GenBank			
B1258(5'-3')		CTGGCGGCCACCAAGTG	
5	Human		
	B: HUMJUNCA	-----	(SEQ ID NO: 62)
	C: HUMJUNA	A-C--T---T-----	(SEQ ID NO: 63)
10	D: HUMJUNDR	A-----TCCAA	(SEQ ID NO: 64)
	Mouse		
	B: MUSJUNBA	-----	(SEQ ID NO: 65)
	C: MUSJUNC	A-T--C---T-----	(SEQ ID NO: 66)
15	MUSCJUN	A-T--C---T-----	(SEQ ID NO: 67)
	MUSJUN	A-T--C---T-----	(SEQ ID NO: 68)
	D: MUSJUND	-----C-----CCCG-	(SEQ ID NO: 69)
	MUSJUNDA	-----T-----GCCA-	(SEQ ID NO: 70)
	MUSJUNDR	-----T-----GCCA-	(SEQ ID NO: 71)
20	Rat		
	C: RATRJG9	-----T-----GCCAA	(SEQ ID NO: 72)
	RATJUNAP	A-C--T---T-----	(SEQ ID NO: 73)
	Chicken		
	C: CHKJUN	-C-----G-----CCC	(SEQ ID NO: 74)
25	Quail		
	C: QULJUN	GGC-----AG----TG-	(SEQ ID NO: 75)
	Drosophila		
	C: DROJUN	G-----T--AT-----	(SEQ ID NO: 76)
30	Homologous sequences in GenBank		
	J04695 Figure 2. Nucleo.	G-----A-----G	(SEQ ID NO: 77)
	M27884 Figure 2. Nucleo.	-----A-----GC	(SEQ ID NO: 78)
	HUMCNPG2 green cone photo.	-----T-AA	(SEQ ID NO: 79)
	HUMCNPR2 green cone photo.	-----T-AA	(SEQ ID NO: 80)
35	HUMPIGMF2 colour-blind pho.	-----T-AA	(SEQ ID NO: 81)

-72-

Table 22. c-jun specific probe (C2147).

	Subtype	GenBank name	C2147.DNA CCACGGCCAACATGCTC	
5	Human			
	B:	HUMJUNCA	-G---A---C---CAGC	(SEQ ID NO: 82)
	C:	HUMJUNA	-----	(SEQ ID NO: 83)
10	D:	HUMJUNDR	-----G-G-C---G	(SEQ ID NO: 84)
	Mouse			
	B:	MUSJUNBA	-G-----C---CAGC	(SEQ ID NO: 85)
	C:	MUSJUN	-----	(SEQ ID NO: 87)
		MUSCJUN	-----	(SEQ ID NO: 88)
		MUSJUNC	-----	(SEQ ID NO: 89)
15	D:	MUSJUND	---C---G-C---G	(SEQ ID NO: 90)
		MUSJUNDR	---C---G-C---G	(SEQ ID NO: 91)
		MUSJUNDA	---C---G-C---G	(SEQ ID NO: 92)
	Rat			
20	C:	RATJUNAP	-----	(SEQ ID NO: 93)
		RATRJG9	---C-----	(SEQ ID NO: 94)
	Chicken			
	C:	CHKJUN	-----	(SEQ ID NO: 95)
	Quail			
	C:	QULJUN	----T-----	(SEQ ID NO: 96)
25	Drosophila			
	C:	DROJUN	CCTACA-----ACC	(SEQ ID NO: 97)
	Homologous sequences in GenBank			
30	MXBPALPA	L.enzymogenes	-----CG--	(SEQ ID NO: 98)
	MXBPALP	L.enzymogenes	-----CG--	(SEQ ID NO: 99)



-73-

Table 23. Human jun-D specific probe (HUMD965)

	Subtype	GenBank name	HUMD965 ACACGCAGGAGCGCATC	
5	Human			
	B:	HUMJUNCA	GG-A--T-----	(SEQ ID NO: 100)
	C:	HUMJUNA	-GT-C-----G---	(SEQ ID NO: 101)
10	D:	HUMJUNDR	-----	(SEQ ID NO: 102)
	Mouse			
	B:	MUSJUNBA	-AGAC-----	(SEQ ID NO: 103)
	C:	MUSJUNC	-GT-T-----G---	(SEQ ID NO: 104)
		MUSCJUN	-GT-T-----G---	(SEQ ID NO: 105)
15		MUSJUN	-GT-T-----G---	(SEQ ID NO: 106)
	D:	MUSJUND	-----A--A-----	(SEQ ID NO: 107)
		MUSJUNDA	-----A--A-----	(SEQ ID NO: 108)
		MUSJUNDR	-----A--A-----	(SEQ ID NO: 109)
	Rat			
20	C:	RATJUNAP	-GT-T-----G---	(SEQ ID NO: 110)
		RATRJG9	-GT-T-----G---	(SEQ ID NO: 111)
	Chicken			
	C:	CHKJUN	-GT-----A-A---	(SEQ ID NO: 112)
	Quail			
25	C:	QULJUN	-GT-----A-A---	(SEQ ID NO: 113)
	Drosophila			
	C:	DROJUN	G--A--T-----	(SEQ ID NO: 114)
	Homologous sequences in GenBank			
	CELPOL11	C.elegans RNA	-----A-AA	(SEQ ID NO: 115)
	ECOPRIAY	E.coli primo.	-ACA-----	(SEQ ID NO: 116)
30	SINOCK82	Ockelbo 82 g.	C-----TGC	(SEQ ID NO: 117)
	TRPPROC	Treponema pa.	G-----TGC	(SEQ ID NO: 118)
	ECOCYSJIHA	E.coli NADPH-	-----T-----G-G	(SEQ ID NO: 119)

-74-

Table 24. Mouse jun-D specific probe (MUSD1063).

	Subtype	GenBank name	MUSD1063.DNA CCCTCAAAAGCCAGAACACCG	
5	Human			
	B:	HUMJUNCA	-G-----GGC-G-----G-G-	(SEQ ID NO: 120)
	C:	HUMJUNA	AA--GC-C-----GC	(SEQ ID NO: 121)
10	D:	HUMJUNDR	-----G--T-----G-	(SEQ ID NO: 122)
	Mouse			
	B:	MUSJUNBA	TA---TCC---TCTG---C	(SEQ ID NO: 123)
	C:	MUSJUNC	AA--GC-T-----GC	(SEQ ID NO: 124)
15		MUSJUN	AA--GC-T-----GC	(SEQ ID NO: 125)
		MUSCJUN	AA--GC-T-----GC	(SEQ ID NO: 126)
	D:	MUSJUNDA	-----	(SEQ ID NO: 127)
		MUSJUNDR	-----	(SEQ ID NO: 128)
		MUSJUND	-----	(SEQ ID NO: 129)
20	Rat			
	C:	RATRJG9	G-----GCC---GC-C---TT	(SEQ ID NO: 130)
		RATJUNAP	G-----GCC---GC-C---TT	(SEQ ID NO: 131)
	Chicken			
	C:	CHKJUN	A---G-GG-A-A-----G---	(SEQ ID NO: 132)
25	Quail			
	C:	QULJUN	---C--G-T-----TG--G-A	(SEQ ID NO: 133)
	Drosophila			
	C:	DROJUN	A-----GGA---TGTGGCGC	(SEQ ID NO: 134)
30	Homologous sequences in GenBank			
	M27221		TG-----T--T-----	(SEQ ID NO: 135)
	DROAMY	D.erecta	G-----G-A--T-----	(SEQ ID NO: 136)
	DROAMYQ	D.erecta	G-----G-A--T-----	(SEQ ID NO: 137)
35	HUMIGCMUDE	Immunoglob.	A--C-----GTT	(SEQ ID NO: 138)

-75-

Table 25. Human-rodent common jun-B specific probes.

	Name in GenBank	B-504 (5'-3')	Seq. ID No:	B-739 (5'-3')	Seq. ID No:	Size (bp)
Mouse		CACGACTACAACTCCTGAAC		GGACAGTACTTTACCCCG		
jun-B 5	MUSJUNBA	-----	600	-----	321	251
c-jun	MUSCJUN	A-AT--A-T-----AT-----A	601	ACC-----T---G-G-----AA	322	
jun-D	MUSJUND	T-----G-C-C-T-----TTCC	602	ACG-----T-C-C-----GAA	323	
Human					324	
jun-B	HUMJUNCA	-----	603	-----	325	251
c-jun 10	HUMJUNA	AGTA--CC--CA-----	604	---A--C-GCGG-----AC	326	
jun-D	HUMJUND	T-----G-C-C-T-----TTTG	605	TC-----T-C-C-----AA	327	
GenBank						
	MUSMET1	-----T---CG-----GTA	606			
	RATMYHCD2			-----C-----AT---A	328	

\*: Size of PCR products using B-504 and B-739 as primers.

\*\*: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

-76-

Table 26. Human-rodent common c-jun specific probes.

	Name in GenBank	C-2101 (5'-3') GAGGAAAAGTGAAACCTTCAAAGC	Seq. ID No:	C-2219 (5'-3') GCCAACTCATGCTAACGCAG	Seq. ID No:	Size (bp)
Mouse						
jun-B 5	MUSJUNBA	-----C--G-----G--AC-C--G--	614	-----GT-GC-----GG-GTC	329	138
c-JUN	MUSCJUN	-----	615	-----	330	
jun-D	MUSJUND	-----G-----C--G-----C-C--AG	616	T-----A-GCC--A-G-----A	331	
Human						
jun-B	HUMJUNCA	-----C--G-----G--GC-C--G--	617	-----G-C-CC-AG-----GC	332	138
c-jun 10	HUMJUNA	-----	618	-----	333	
jun-D	HUMJUNDR	--A--G-----G-----C-C--GAG	619	T-----G-GCC--A-G-----A	334	
GenBank	HUMGASTA MUSOCT22	A-A-----AA-----A--A-----	620	C-----G-----GGC	335	

\*: Size of PCR products using C-2101 and C-2219 as primers.

\*\*: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

SUBSTITUTE SHEET

-77-

Table 27. Human-rodent common jun-D specific probes.

	Name in GenBank	D-916 (5'-3')	Seq. ID No:	D-1153 (5'-3')	Seq. ID No:	Size (bp)
Mouse				AAAGTCCTCAGCCACGTCAC		
jun-B 5	MUSJUNBA	CTG-G-----ACT--	628	--G---A-G-C---T-----G-	336	
c-jun	MUSCJUN	T-A-----AG---AGAC	629	-----A-G-A-----T---	337	258
jun-D	MUSJUND	-----	630	-----	338	
Human						
jun-B	HUMJUNCA	C---C-----CC-----C	631	--G---A-G-C-----G-	339	
C-jun 10	HUMJUNA	T-A---TT-----CG--	632	-----A-G-A-----T---	340	
jun-D	HUMJUNDR	-----	633	-----	341	258
GenBank						
	HUMTRKR	-----TG-----	634	---A-----A---CTG	342	
	HUMBSSL					

\*: Size of PCR products using D-916 and D-1153 as primers.

\*\*: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

### Identifying Probes and PCR Primers

Probes and PCR primers for use in the methods of the present invention can be identified in any way known to the art. Preferably, however, such probes and primers are identified by a computer. We have developed a novel computer system for identifying the sequences to be used in such probes and primers. This system is an automated system which allows the user to calculate and design extremely accurate oligonucleotide probes and PCR primers.

The software of the present invention runs under Microsoft Windows® on IBM® compatible personal computers (PC's). This invention allows a researcher to design oligonucleotide probes based on the GenBank database of DNA and mRNA sequences. The present invention further allows examination of probes for specificity or commonality with respect to a user-selected target gene sequences. Hybridization strength between a probe and a target subsequence of DNA or mRNA can be estimated through a hybridization strength model. Quantitatively, hybridization strength is given as the melting temperature ( $T_m$ ).

Two models for estimating hybridization strength models are supported by this invention: 1) the Mismatch Model and 2) the H-Site Model. In either case, the user can select the following calculations for each probe, results of which are then made available for display and analysis: 1) Sequence, Melting Temperature ( $T_m$ ) and Hairpin characteristics (a hairpin is a nucleotide sequence that is homologous to itself and can "fold back" with one portion of the probe hybridizing to another portion of the same probe); 2) Hybridization to other species within the preparation mixture; and (3) Location and  $T_m$  for the strongest hybridizations. The results of the invention's calculations are then displayed on a Mitsubishi Probe Selection Diagram (MPSD) which is a graphic display of all potential hybridizations between the target mRNA and the probe sequences in the preparation.

The Main OligoProbe Design Station dialog window controls all user-definable settings in the program. The user is offered a number of options at this window. The File option allows the user to print, print in color, save selected probes, and exit the program. The Preparation option allows the user to open and create preparation (PRP) files. The Models option allows the user to chose between the two hybridization models currently supported by the OligoProbe DesignStation: 1) the H-Site Model and 2) the Mismatch Model.

If the user selects the H-Site Model option, the melting temperature for each probe and the nucleation threshold parameters can be set. The nucleation threshold is the number of base pairs constituting a nucleation site (a subsequence with an exact match). If the user selects the Mismatch Model option, the probe length and mismatches (N) can be set.

#### Mismatch Model

The Mismatch Model is used for designing DNA and mRNA probes utilizing sequence database information from sources such as GenBank. In this Model, hybridization strength is related only to the number of base pair mismatches between a probe and its target. Generally, the more mismatches a user allows when setting parameters, the more probes will be identified. The Mismatch Model does not take into account the GC content of candidate probes so there is no calculation of the probe's binding strength.

The basic technologies employed by the Mismatch model are hashing and continuous seed filtration. Hashing involves the application of an algorithm to the records in a set of data to obtain a symmetric grouping of the records. When using an indexed set of data such as a database, hashing is the process of transforming a record key to an index value for storing and retrieving a record. The Mismatch Model is essentially a quick process for determining exact and inexact matching between DNA and mRNA sequences to support the Mitsuhashi Probe Selection Diagram (MPSD).

The algorithm used by the Mismatch Model is based on the Waterman-Pevzner Algorithm (WPALG), which is a computer-based probe selection process. Essentially, this is a combination of new and improved pattern matching processes. See Hume and Sunday (1991, Ref. 4), Landau *et al* (1986-1990, Refs. 6, 7, 8), Grossi and Luccio (1989, Ref. 3), and Ukkonen (1982, Ref. 14).

There are three principal programs that make up the Mismatch Model in this implementation of the invention. The first is designated by the inventors as "k\_diff." WPALG uses k\_diff to find all locations of matches of length greater than or equal to one (1) (length is user-specified) with less than or equal to k number of mismatches (k is also user-specified) between the two sequences. If a candidate oligonucleotide probe fails to match that well, it is considered unique. k\_diff uses hashing and continuous seed filtration, and looks for homologs by searching GenBank and other databases with similar file formats. The technique of continuous seed filtration allows for much more efficient searching than previously implemented techniques.

A seed is defined in this invention to be a subsequence having a length equal to the longest exact match in the worst case scenario. For example, suppose the user selects a probe length (l) of 18, with 2 or fewer mismatches (k). If a match exists with 2 mismatches, then there must be a perfectly matching subsequence of length equal to 6.

5 Once the seed length has been determined, the Mismatch Model looks at all substrings of that seed length (in this example, the seed length would be 6), finds the perfectly matched base pair subsequence of length equals 6, and then looks to see if this subsequence extends to a sequence of length equal to the user selected probe length (i.e., 18 in this example). If so, a candidate probe has been found that meets the user's criteria.

10 Where the seed size is large (i.e., a long string of unique nucleotides), the program allocates a relatively large amount of memory for the hash table. This invention has an option that allows memory allocation for GenBank entries just once at the beginning of the program, instead of reallocating memory for each GenBank entry. This reduces input time for GenBank entries by as much as a factor of two (2), but this method requires the

15 user to know the maximum GenBank entry size in advance.

A probe is found to hybridize if it has k or fewer mismatches with a target sequence from the database or file searched. The hit extension time for all appropriate parameters of the Mismatch Model has been found by experimentation to be less than thirty-five (35) seconds, except in one case where the minimum probe length (l) was set to 24 and the

20 maximum number of mismatches (k) was set to four (4). This situation would rarely be used in real gene localization experiments because the hybridization conditions are too weak.

#### H-Site Model

In this embodiment of the invention, the second hybridization strength model is termed the H-Site Model. One aspect of the H-Site Model uses a generalization of an

25 experimental formula to analyze nucleotide binding strength. The basic formula on which this aspect of the model is built is as follows:

$$T_m = 81.5 - 16.6(\log[\text{Na}]) - .63 \%(\text{formamide}) + .41 (\%(\text{G} + \text{C})) - 600 / N$$

In this formula, log[Na] is the log of the sodium concentration. %(G + C) is the fraction

30 of matched base pairs which are G-C complementary, and N is the probe length. This formula relates the fact that melting temperature is a function of both probe length and percent GC content. This basic formula has been modified in this invention to account for the presence of mismatches. Each percent of mismatch reduces the melting temperature by an average of 1.25° (2°C for an AT mismatch, and 4°C for a GC



mismatch). This formula is, however, an approximation. The actual melting temperature might potentially differ from this approximation, especially for short probes or probes with a relatively large number of mismatches.

Hybridization strength in the H-Site Model is related to each of the following factors:  
5 1) "binding region"; 2) type of mismatch (GC or AT substitution); 3) length of the probe;  
4) GC content of the binding region; and 5) existence of a "nucleation site" (a subsequence with an exact match). The type of mismatch and GC content of the binding region from each sequence contributes to a candidate probe's binding strength. The binding strength from each probe is thereby determined enabling the user to select an optimal probe.

10 The fundamental assumption of the H-Site Model is that binding strength is mostly determined by a paired subsequence of the probe and target, called the binding region. If the subsequence binding region contains more GC pairs than AT pairs, the binding strength will be higher due to the greater number of hydrogen bonds between G and C bases (three bonds) in comparison to A and T bases (two bonds). Thus, GC rich probes  
15 have a higher melting temperature and subsequently form stronger hybridizations.

In the H-Site Model the program determines optimal probes, ideally without any mismatches to the target gene. With this model, however, a candidate probe can have more AT mismatches if the sequence is GC rich. The amount of allowable AT mismatches in a specific sequence is determined in the present invention program by looking primarily  
20 at subsequence regions of the probe and target that match without penalizing the probe for areas that mismatch. If the mismatches are located at either or both of the ends of the binding region, there is little effect on the overall stability of the base-pairing. Centrally located mismatches in the binding region are much more deleterious, as this will significantly lower the binding strength of the probe.

25 The formula cited above for the melting temperature applies within the binding region. The length of the probe is used to calculate percentages, but all other parameters of the formula are applied to the binding region only. The H-Site Model further assumes the existence of a nucleation site. The length of this nucleation site may be set by the user. Typically, a value of 8 to 10 base pairs is used. To complete the H-Site Model, the  
30 binding region is chosen so as to maximize the melting temperature  $T_m$  among all regions containing a nucleation site, assuming one exists (otherwise,  $T_m = 0$ ).

The H-Site Model is more complex than the Mismatch Model discussed above in that hybridization strength is modeled as a sum of multiple subsequence contributions, with matches generally providing positive binding energy and mismatches generally providing

-82-

negative binding energy. The exact binding energies to be used depend only on the matched or mismatched pair. These coefficients may be specified by the user, although in the current version of this invention these coefficients are not explicitly user-selectable, but rather are selected to best fit the hybridization strength formulas developed by Itakura  
5 et al (1984, Ref. 5), Bolton and McCarthy (1962, Ref. 2), Benner et al (1973, Ref. 1), and Southern (1975, Ref. 13).

A unique aspect of the H-Site Model is that hybridization strength is determined by the optimal binding region between the candidate probe and binding locus. This binding region is called the hybridization site, or h-site, and is selected so as to maximize overall  
10 hybridization strength, so that mismatches outside the binding region do not detract from the estimated hybridization strength. Several other unique features of the H-Site Model include the fact that it is more oriented toward RNA and especially cDNA sequences than DNA sequences, and the fact that the user has control over preparation and environmental variables.

15 The emphasis on RNA and cDNA sequences allows the user to concentrate on coding regions of genes, rather than necessitating sorting through all of a genomic sequence for the desired probe. The enhanced user control over environmental and preparation variables allows the user to more accurately simulate laboratory conditions that closely correspond with any experiments he or she is conducting. Further, this implementation  
20 of the invention does some preliminary preprocessing of the GenBank database to sort out and select the cDNA sequences. This is done by locating a keyword (in this case CDS) in each GenBank record, thereby eliminating any sequences containing introns.

The Mitsuhashi Probe Selection Diagram (MPSD), Figures 28A and 28B, is a key feature of this invention, as it is a unique way of visualizing the results of the probe  
25 designed by the Mismatch and H-Site Models. It is a graphic display of all of the hybridizations of candidate oligonucleotide probes and the target with all sequences in the preparation. Given a gene sequence database and a target mRNA sequence, the MPSD graphically displays all of the candidate probes and their hybridization strengths with all sequences from the database. In the present implementation, each melting temperature  
30 is displayed as a different color, from red (highest T<sub>m</sub>) to blue (lowest T<sub>m</sub>). The MPSD allows the user to see visually the number of false hybridizations at various temperatures for all candidate probes, and the sources of these false hybridizations (with a loci and sequence comparison). A locus may be a specific site or place, or, in the genetic sense,

a locus is any of the homologous parts of a pair of chromosomes that may be occupied by allelic genes.

These probes may then be used to test for the presence of precursors of specific proteins in living tissues. The oligonucleotide probes designed with this invention may be used for medical diagnostic kits, DNA identification, and potentially continuous monitoring of metabolic processes in human beings. The present implementation of this computerized design tool runs under Microsoft® Windows™ v. 3.1 (made by Microsoft Corporation; Redmond, Washington) on IBM® compatible personal computers (PC's).

The H-Site Model of this invention is unique in that it offers a multitude of information on selected probes and original and distinctive means of visualizing, analyzing and selecting among candidate probes designed with the invention. Candidate probes are analyzed using the H-Site Model for their binding specificity relative to some known set of mRNA or DNA sequences, collected in a database such as the GenBank database. The first step involves selection of candidate probes at some or all the positions along a given target. Next, a melting temperature model is selected, and an accounting is made of how many false hybridizations each candidate probe will produce and what the melting temperature of each will be. Lastly, the results are presented to the researcher along with a unique set of tools for visualizing, analyzing and selecting among the candidate probes.

This invention is both much faster and much more accurate than the methods that are currently in use. It is unique because it is the only method that can find not only the most specific and unique sequence, but also the common sequences. Further, it allows the user to perform many types of analysis on the candidate probes, in addition to comparing those probes in various ways to the target sequences and to each other.

Therefore, it is the object of this invention to provide a practical and user-friendly system that allows a researcher to design both specific and common oligonucleotide probes, and to do this in less time and with much more accuracy than currently done.

This invention is employed in the form best seen in Figure 20. There, the combination of this invention consists of an IBM® compatible personal computer (PC), running software specific to this invention, and having access to a distributed database with the file formats found in the GenBank database and other related databases.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred

methods and materials are now described. All publications mentioned hereunder are incorporated herein by reference.

The preferred computer hardware capable of operating this invention involves of a system with at least the following specifications (Figure 20): 1) an IBM® compatible PC, generally designated 1A, 1B, and 1C, with an 80486 coprocessor, running at 33 Mhz or faster; 2) 8 or more MB of RAM, 1A; 3) a hard disk 1B with at least 200 MB of storage space, but preferably 1 GB; 4) a VGA color monitor (1C) with graphics capabilities of a size sufficient to display the invention's output in readable format, preferably with a resolution of 1024 x 768; and 5) a 580 MB CD ROM drive 5 (1B of Figure 20 generally refers to the internal storage systems included in this PC, clockwise from upper right, two floppy drives, and a hard disk). Because the software of this invention preferably has a Microsoft® Windows™ interface, the user will also need a mouse 2, or some other type of pointing device.

The preferred embodiment of this invention would also include a laser printer 3 and/or a color plotter 4. The invention may also require a modem (which can be internal or external) if the user does not have access to the CD ROM versions of the GenBank database 8 (containing a variable number of gene sequences 6). If a modem is used, information and instructions are transmitted via telephone lines to and from the GenBank database 8. If a CD ROM drive 5 is used, the GenBank database (or specific portions of it) is stored on a number of CDs.

The computer system should preferably have at least the Microsoft® DOS version 5.0 operating system running Microsoft® Windows™ version 3.1. All of the programs in the preferred embodiment of the invention were written in the Borland® C++ (Borland International, Inc.; Scotts Valley, CA) computer language. It should be noted that subsequently developed computers, storage systems, and languages may be adapted to utilize this invention and vice versa.

This inventive computer program is designed to enable the user to access DNA, mRNA and cDNA sequences stored either in the GenBank or in databases with similar file formats. GenBank is a distributed flat file database made up of records, each record containing a variable number of fields in ASCII file format. The stored database itself is distributed, and there is no one database management system (DBMS) common to even a majority of its users. One general format, called the line type format, is used both for the distributed database and for all of GenBank's internal record keeping. All data and system files and indexes for GenBank are kept in text files in this line type format.

The primary GenBank database is currently distributed in a multitude of files or divisions, each of which represents the genome of a particular species (or at least as much of it as is currently known and sequenced and publicly available). The GenBank provides a collection of nucleotide sequences as well as relevant bibliographic and biological annotation. Release 72.0 (6/92) of the GenBank CD distribution contains over 71,000 loci with a total of over ninety-two (92) million nucleotides. GenBank is distributed by IntelliGenetics, of Mountain View, CA, in cooperation with the National Center for Biotechnology Information, National Library of Medicine, in Bethesda, MD.

1. Overall Description of the OligoProbe DesignStation

a. **General Theory**

The intent of this invention is to provide one or more fast processes for performing exact and inexact matching between DNA sequences to support the Mitsuhashi Probe Selection Diagram (MPSD) discussed below, and other analysis with interactive graphical analysis tools. Hybridization strength between a candidate oligonucleotide probe and a target subsequence of DNA, mRNA or cDNA can be estimated through a hybridization strength model. Quantitatively, hybridization strength is given as the melting temperature ( $T_m$ ). Currently, two hybridization strength models are supported by the invention: 1) the Mismatch Model and 2) the H-Site Model.

b. **Inputs**

i. Main OligoProbe DesignStation Dialog Window

The Main OligoProbe DesignStation dialog window, Figure 21, controls all user-definable settings. This window has a menu bar offering five options: 1) File 10; 2) Preparation 80; 3) Models 70; 4) Experiment 90; and 5) Help 50. The File 10 option allows the user to print, print in color, save selected probes, and exit the program. The Preparation 80 option allows the user to open and create preparation (PRP) files.

The Models 70 option allows the user to choose between the two hybridization models currently supported by the OligoProbe DesignStation: 1) the H-Site Model 71 and 2) the Mismatch Model 75. If the user selects the H-Site Model 71 option, the left hand menu of Figure 21C is displayed and the user sets the following model parameters: 1) the melting temperature  $T_m$  72 for which probes are being designed (i.e., the melting temperature that corresponds to a particular experiment or condition the user desires to simulate); and 2) the nucleation threshold 73, which is the number of base pairs constituting a nucleation site. If the user selects the Mismatch Model 75 option, the right

hand menu of Figure 21C is displayed and the user sets the following model parameters:  
1) probe length 76, which is the number of base pairs in probes to be considered; and 2)  
mismatch N 77, which is the maximum number of mismatches constituting a hybridization.  
Computation of the user's request takes longer with the H-Site Model if the threshold 73  
5 setting is decreased, but longer with the Mismatch Model if the number of mismatches K  
77 is increased.

In addition, for both Model options the user chooses the target species 11 DNA or  
mRNA for which probes are being designed and the preparation 12, a file of all sequences  
with which hybridizations are to be calculated. A sample of a target species file is shown  
10 in Figure 47 (humbjunx.cds), while a sample of a preparation file is shown in Figures 48A-  
48C (junmix.seq). Each of these inputs is represented by a file name and extension in  
standard DOS format. In the target species and preparation fields, the file format follows  
the GenBank format with each of the fields having a default file extension. Pressing the  
"OK" button 91 (Figure 21C) will initiate processing, while pressing the "Cancel" button 93  
15 will stop the processing.

The Experiment 90 option and the Help 50 option are expansion options not yet  
available in the current implementation of the invention.

c. **Processing**

Figures 22A and 22B are flow charts of the overall OligoProbe DesignStation  
20 Program, illustrating its sequence and structure. Generally, the main or "control" program  
of the OligoProbe DesignStation performs overall maintenance and control functions. This  
program, as illustrated in Figures 22A and 22B, accomplishes the general housekeeping  
functions 51, such as defining global variables. The user-friendly interface 53, carries out  
the user-input procedures 55, the file 57 or database 59 access procedures, calling of the  
25 model program 62 or 63 selected by the user, and the user-selected report 65 or display  
67, 69, 71 and 73 features. Each of these features is discussed in more detail in later  
sections, with the exception of the input procedures, which involves capturing the user's  
set-up and control inputs.

d. **Outputs**

i. The Mitsuhashi Probe Selection Diagram Window

The Mitsuhashi Probe Selection Diagram (MPSD), Figure 23, is a key feature of the  
invention as it is a unique way of visualizing the results of the program's calculations. It  
is a graphic display of all of the hybridizations of probes with the target oligonucleotides  
in the preparation. Specifically, given a nucleotide sequence database and a target mRNA,



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

<b>(51) Classification internationale des brevets <sup>6</sup> :</b> <b>C12N</b>	<b>A2</b>	<b>(11) Numéro de publication internationale:</b> <b>WO 97/28247</b> <b>(43) Date de publication internationale:</b> 7 août 1997 (07.08.97)
<b>(21) Numéro de la demande internationale:</b> PCT/FR97/00167 <b>(22) Date de dépôt international:</b> 28 janvier 1997 (28.01.97) <b>(30) Données relatives à la priorité:</b> 96/01004 29 janvier 1996 (29.01.96) <b>FR</b> <b>(71) Déposant (pour tous les Etats désignés sauf US):</b> BIOCEM [FR/FR]; Campus Universitaire des Cézéaux, 24, avenue des Landais, F-63170 Aubière (FR). <b>(72) Inventeurs; et</b> <b>(75) Inventeurs/Déposants (US seulement):</b> LUDEVID, Dolorès [ES/ES]; Centro de Investigacion y Desarrollo CSIC, Jordi Girona Salgado, 18-26, E-08034 Barcelone (ES). TORRENT, Margarita [ES/ES]; Centro de Investigacion y Desarrollo CSIC, Jordi Girona Salgado, 18-26, E-08034 Barcelone (ES). ALVAREZ, Inaki [ES/ES]; Centro de Investigacion y Desarrollo CSIC, Jordi Girona Salgado, 18-26, E-08034 Barcelone (ES). PEREZ, Pascual [FR/FR]; 21, allée du Parc, F-63110 Beaumont (FR). <b>(74) Mandataires:</b> GUTMANN, Ernest etc.; Ernest Gutmann - Yves Plasseraud S.A., 3, rue Chauveau-Lagarde, F-75008 Paris (FR).		<b>(81) Etats désignés:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, brevet ARIPO (KE, LS, MW, SD, SZ, UG), brevet eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Publiée</b> <i>Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.</i>
<b>(54) Title:</b> AMINO ACID-ENRICHED PLANT PROTEIN RESERVES, PARTICULARLY LYSINE-ENRICHED MAIZE $\gamma$ -ZEIN, AND PLANTS EXPRESSING SUCH PROTEINS <b>(54) Titre:</b> PROTEINES DE RESERVE DE PLANTES ENRICHIES EN ACIDES AMINES, NOTAMMENT $\gamma$ -ZEINE DE MAIS ENRICHIE EN LYSINE, PLANTES EXPRIMANT CES PROTEINES <b>(57) Abstract</b> <p>An oligonucleotide including at least one concatenation coding for a polypeptide of formula (P-K)<sub>n</sub>, wherein n is an integer of at least 2, P is a proline amino acid residue, K is a lysine amino acid residue, and the sign "-" is a bond, particularly a peptide bond, between the two amino acid residues. The n (P-K) units are also bound together by such bonds, e.g. peptide bonds.</p> <b>(57) Abrégé</b> <p>L'invention a pour objet un oligonucléotide comprenant au moins un enchaînement codant pour un polypeptide répondant à la formule (P-K)<sub>n</sub>, dans laquelle: n est un entier supérieur ou égal à 2, P représente un résidu d'acide aminé Proline, K représente un résidu d'acide aminé Lysine, le signe "-" symbolise une liaison entre les deux résidus d'acides aminés et notamment une liaison de type peptidique, les n unités (P-K) étant liées entre elles également par de telles liaisons par exemple de type peptidique.</p>		





# **UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AT	Arménie	GB	Royaume-Uni	MW	Malawi
AT	Autriche	GE	Géorgie	MX	Mexique
AU	Australie	GN	Guinée	NE	Niger
BB	Barbade	GR	Grèce	NL	Pays-Bas
BE	Belgique	HU	Hongrie	NO	Norvège
BF	Burkina Faso	IE	Irlande	NZ	Nouvelle-Zélande
BG	Bulgarie	IT	Italie	PL	Pologne
BJ	Bénin	JP	Japon	PT	Portugal
BR	Brésil	KE	Kenya	RO	Roumanie
BY	Bélarus	KG	Kirghizistan	RU	Fédération de Russie
CA	Canada	KP	République populaire démocratique de Corée	SD	Soudan
CF	République centrafricaine	KR	République de Corée	SE	Suède
CG	Congo	KZ	Kazakhstan	SG	Singapour
CH	Suisse	LI	Liechtenstein	SI	Slovénie
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovaquie
CM	Cameroon	LR	Libéria	SN	Sénégal
CN	Chine	LT	Lituanie	SZ	Swaziland
CS	Tchécoslovaquie	LU	Luxembourg	TD	Tchad
CZ	République tchèque	LV	Lettonie	TG	Togo
DE	Allemagne	MC	Monaco	TJ	Tadjikistan
DK	Danemark	MD	République de Moldova	TT	Trinité-et-Tobago
EE	Estonie	MG	Madagascar	UA	Ukraine
ES	Espagne	ML	Mali	UG	Ouganda
FI	Finlande	MN	Mongolie	US	Etats-Unis d'Amérique
FR	France	MR	Mauritanie	UZ	Ouzbékistan
GA	Gabon			VN	Viet Nam



Protéines de réserve de plantes enrichies en acides aminés, notamment  $\gamma$ -zéine de maïs enrichi en lysine Plantes exprimant ces protéines

5           La présente demande a pour objet de nouveaux moyens permettant la préparation de plantes exprimant des protéines de réserve enrichies en acides aminés déficitaires dans les protéines de réserve normales, et en particulier de protéines de réserves enrichies en lysine. L'invention a également pour objet les protéines de réserve ainsi modifiées, ainsi que  
10 des plantes exprimant ces protéines de réserve modifiées.

          De nombreuses plantes, le cas échéant après transformation par des étapes physico-chimiques, sont d'une importance économique majeure dans l'alimentation humaine ou animale, et le problème de l'amélioration de leur qualité nutritionnelle a déjà donné lieu à différents types de  
15 recherches. En particulier, pour remédier à l'insuffisante présence de certains acides aminés dans les protéines de réserve des plantes, on a mis au point des variétés sélectionnées présentant des qualités nutritionnelles supérieures, ou proposé différentes modifications faisant appel à l'utilisation des techniques du génie génétique afin de favoriser ou  
20 d'augmenter la production dans ces plantes notamment de certains acides aminés déficitaires, et néanmoins importants pour les qualités nutritionnelles de la plante. Ces acides aminés déficitaires sont par exemple la lysine, la méthionine.

          Dans le cadre de la présente demande, les inventeurs se sont  
25 proposé d'apporter une solution originale au problème de l'amélioration des plantes, notamment de l'amélioration de leurs qualités nutritionnelles et ont à cet effet travaillé en premier lieu à partir d'une plante d'importance économique considérable, le maïs. Plus précisément, ils se sont intéressés aux protéines de réserve de l'endosperme des semences de maïs,



protéines qui comprennent notamment les zéines et, en particulier, la  $\gamma$ -zéine.

Au cours du développement des semences de maïs, les cellules de l'endosperme synthétisent de grandes quantités de protéines de réserve, et notamment de  $\alpha$ -,  $\beta$ -, et  $\gamma$ -zéines. Ces zéines sont accumulées dans les corps protéiques dérivés du réticulum endoplasmique (ER).

D'une façon générale, les zéines représentent un groupe complexe de protéines, réparties en plusieurs groupes,  $\alpha$ -,  $\beta$ -,  $\gamma$ -, et  $\delta$ -zéines (Larkins et al, 1989), codées par une famille multigénique (Hagen et Rubenstein, 1980, Gene 13, 239-249). Bien qu'elles présentent une structure variable, ces protéines ont des caractéristiques communes : la présence dans leur structure primaire de répétitions en tandem riches en résidus d'acides aminés de type Proline, la présence de nombreux résidus hydrophobes qui confèrent leur insolubilité à ces protéines dans un milieu aqueux, et l'absence de résidus lysine, acides aminés essentiels pour l'homme et les animaux monogastriques. L'absence de lysine dans toutes les protéines majoritaires (détectées en grande quantité dans l'endosperme) naturellement produites de la famille des zéines conduit à une composition en acides aminés déséquilibrée dans les semences de maïs.

Parmi ces protéines, la  $\gamma$ -zéine de maïs est une protéine ayant un poids moléculaire de 28 kDa, dont la séquence codante a été décrite sous la forme de l'ADNc par Prat et al (Nucleic Acids Research, vol. 13, n° 5, 1985, p. 1494-1504). La séquence complète du gène codant pour la  $\gamma$ -zéine, y compris les séquences non codantes en amont et en aval contenant les éléments de régulation de l'expression, a été décrite par Reina, M. et al (Nucleic Acids Research, vol. 18, n° 21, 1990, p 6426).

Jusqu'à présent, différentes approches ont été envisagées pour augmenter la teneur en lysine des protéines de la famille des zéines. A cet égard, des approches génétiques et moléculaires ont été mises en oeuvre.



Par exemple, des mutants permettant l'obtention de maïs riche en lysine, tels que le mutant opaque-2 (o2) et le mutant floury-2 (fl-2) (Mertz et al, 1964, Science 145, 279-280, Nelson et al, 1965, Science 150, 1469-1470) ont ainsi été proposés et des tentatives pour remédier aux effets néfastes de l'absence de certaines classes de zéïnes, notamment les  $\alpha$ -zéïnes, sur les caractéristiques phénotypiques, par sélection de maïs contenant des gènes modificateurs o2 (Paez et al, 1969, Plant Sci. 9, 251-252, Geetha et al, 1991, Plant Cell 3, 1207-1219) ont été faites.

Une autre approche a consisté à réaliser une action indirecte sur la production de lysine libre, en particulier dans les plantes dicotylédones. Cette technique a été mise en oeuvre en agissant sur la dérégulation des enzymes clé (DHTPS et AK) impliquées dans le cycle de la biosynthèse de la lysine via l'aspartate. Un accroissement sensible des niveaux de lysine libre a été obtenu dans les feuilles, mais pas dans les semences, dans des expériences de transformation de plantes de tabac avec des bactéries E. coli contenant les gènes dapA et des bactéries E. coli contenant le gène lysC (Shaul et Galili, 1992, Plant J. 2, 203-209 et 1993, Plant Mol. Biol. 23, 759-768 ; Perl, A., Schaul O., Galili, G., 1992, Plant Molecular Biology, 19, p. 815-823). Récemment, les mêmes gènes dapA de Corynebacterium et lysC de E. coli ont été utilisés et exprimés sous le contrôle d'un promoteur spécifique des semences dans des plantes de soja. L'expression de ces deux enzymes dans le soja a conduit à une augmentation de cinq fois de la teneur en lysine dans les semences (Falco et al, 1995, BIO-Technology 13, 577-582).

D'autres auteurs (Wallace et al, 1988, Science 240, 662-664) ont tenté d'augmenter la lysine dans l' $\alpha$ -zéïne (19 kDa) des semences de maïs en incorporant de façon ponctuelle des résidus lysine en différentes positions dans la molécule d' $\alpha$ -zéïne. L'expression de ces constructions dans des oocytes de Xenopus a conduit à l'assemblage correct des zéïnes riches en lysine dans les vésicules analogues des corps protéiques.





Cependant, l' $\alpha$ -zéine normale et la forme modifiée enrichie en lysine étaient dégradées lorsqu'elles étaient exprimées dans des semences de tabac (Othani et al, 1991, Plant Mol. Biol. 16:117).

On ne disposait donc pas à ce jour de connaissances quant aux  
5    moyens pouvant permettre l'expression d'une zéine enrichie en lysine dans les cellules la produisant naturellement dans le maïs, c'est-à-dire dans les cellules de l'endosperme. A fortiori l'expression dans d'autres cellules végétales de zéines enrichies en lysine n'était pas maîtrisée.

L'un des buts de l'invention est donc de proposer des moyens  
10    permettant d'obtenir une zéine enrichie en lysine, en particulier une  $\gamma$ -zéine de maïs enrichie en lysine, cette protéine étant exprimée notamment dans les cellules des semences de maïs et en particulier dans les cellules de l'endosperme, ladite protéine modifiée étant en outre exprimée de façon telle que ses propriétés quant à la localisation et l'accumulation au niveau  
15    du réticulum endoplasmique et des corps protéiques dérivés soient préservées.

L'expression « enrichie en lysine » signifie, dans le contexte de la présente demande, que la protéine comporte un nombre augmenté de résidus lysine par rapport à la protéine naturelle dont elle est dérivée, par  
20    exemple par suite d'une modification de la séquence de nucléotides qui l'exprime.

L'invention propose aussi des moyens pour obtenir l'expression de protéines, de préférence de  $\gamma$ -zéines enrichies en lysine, dans des cellules végétales de différents tissus, par exemples des tissus de feuilles ou de  
25    racines, et ce le cas échéant dans des cellules végétales de plantes n'exprimant pas naturellement la protéine, en particulier la  $\gamma$ -zéine dont la production est recherchée.



Par ailleurs, selon un mode de réalisation particulier de l'invention, d'autres protéines de réserve peuvent être enrichies en lysine selon des modalités analogues.

Les inventeurs se sont proposé dans un premier temps, pour  
5 réaliser l'invention, d'introduire dans le gène codant pour la  $\gamma$ -zéine ou pour d'autres protéines de réserve du maïs ou d'autres plantes, ou dans la séquence codante de ce gène, des séquences codant pour des polypeptides enrichis en lysine, afin d'obtenir la production de  $\gamma$ -zéines ou d'autres protéines enrichies en lysine et donc, de semences enrichies en  
10 lysine. Différents sites de la séquence codante du gène de la  $\gamma$ -zéine ont été identifiés comme sites permissifs, (appelés encore "sites neutres") pour préparer des séquences de nucléotides ainsi modifiées.

La présente demande propose donc des moyens pour transformer le gène codant pour la  $\gamma$ -zéine de maïs ou pour transformer toute séquence  
15 de nucléotides codant pour la  $\gamma$ -zéine et dérivée de ce gène, de façon à obtenir à partir de l'expression du gène modifié ou plus généralement de la séquence de nucléotides modifiée, une protéine enrichie en lysine; ces moyens comprennent en particulier des oligonucléotides synthétiques codant pour une séquence d'acides aminés comportant des résidus lysine.

20 L'invention a également pour objet des séquences de nucléotides recombinantes ou séquences chimères susceptibles de coder pour une  $\gamma$ -zéine enrichie en lysine.

Entrent également dans le cadre de l'invention des cellules hôtes transformées par de telles séquences, et en particulier des cellules  
25 végétales, par exemple des cellules permettant la régénération de plantes, ainsi que des plantes ou des parties de plantes (tissus, organes,...) contenant de telles cellules et produisant de façon stable des protéines de réserve modifiées, et en particulier des  $\gamma$ -zéines enrichies en lysine.



L'invention a également pour objet lesdites protéines modifiées, par exemple enrichies en lysine, ainsi que des anticorps dirigés contre ces protéines.

Un oligonucléotide approprié pour la réalisation de l'invention, utilisable pour la préparation de séquences de nucléotides recombinantes, est caractérisé en ce qu'il comprend au moins un enchaînement codant pour un polypeptide répondant à la formule  $(P-K)_n$ , dans laquelle :

- n est un entier supérieur ou égal à 2,
- P représente un résidu d'acide aminé Proline,
- 10      - K représente un résidu d'acide aminé Lysine,
- le signe "-" symbolise une liaison entre les deux résidus d'acides aminés et notamment une liaison de type peptidique, les n unités (P-K) étant liées entre elles également par de telles liaisons par exemple de type peptidique.

15      Un oligonucléotide selon l'invention est donc caractérisé selon un premier mode de réalisation, en ce qu'il comprend une séquence codant pour une suite de motifs répétés comportant deux acides aminés.

Les codons de l'oligonucléotide peuvent être identiques pour tous les résidus Proline et/ou pour tous les résidus lysine. Ils peuvent aussi être  
20      différents pour un même résidu d'acide aminé, la variation tenant compte de la dégénérescence du code génétique.

De préférence, cet oligonucléotide est formé par une séquence codant pour plus de 2 unités (P-K). De préférence n est inférieur ou égal à 30, en particulier inférieur à 20 et avantageusement n est égal à 4, 5, 6, 7,  
25      8, 9 ou 10 ou 15.

Les "oligonucléotides" selon l'invention peuvent être synthétisés chimiquement par toute technique disponible.

Le terme "polypeptide" par référence à l'enchaînement  $(P-K)_n$  désigne, dans le cadre de la présente demande, une séquence d'acides



aminés comportant plus de 2 résidus d'acides aminés et pouvant comporter jusqu'à 60 résidus d'acides aminés.

Selon une première variante de réalisation de l'invention, l'oligonucléotide comprend plusieurs enchaînements codant pour un polypeptide répondant à la formule  $(P-K)_n$ , identiques ou différents, associés en tandem.

Ces oligonucléotides sont soit des répétitions d'un même enchaînement, soit des associations d'enchaînement différents. Le nombre d'enchaînements associés est variable, pouvant par exemple être compris entre 2 et 10 enchaînements.

Selon une autre variante de réalisation de l'invention, l'oligonucléotide précédemment défini est caractérisé en ce qu'il comprend au moins un enchaînement codant pour un polypeptide répondant à la formule  $(P-K)_n$ , dans laquelle la séquence des  $n$  unités  $(P-K)$  est interrompue par un ou plusieurs résidus d'acides aminés différents des résidus P ou K.

De préférence, les acides aminés supplémentaires incorporés dans la séquence formée par les unités  $(P-K)$  sont choisis pour ne pas modifier l'organisation du polypeptide codé par l'oligonucléotide, ou à tout le moins pour ne pas provoquer d'interaction avec des acides aminés d'une protéine dans laquelle ledit polypeptide serait incorporé, dans des conditions qui affecteraient la structure et/ou la fonction et/ou la localisation de cette protéine.

Ceci peut en particulier être le cas lorsque le nombre d'unités  $(P-K)$  est important ou que plusieurs enchaînements formés de séquences codant pour des motifs  $(P-K)_n$  sont associés en tandem et que la préparation de l'oligonucléotide correspondant nécessite que soient synthétisées plusieurs séquences nucléotidiques qui sont ensuite associées par exemple au moyen de séquences de liaison (linkers).





Selon un autre mode de réalisation de l'invention, l'oligonucléotide est tel que l'enchaînement codant pour le polypeptide comprenant les n unités (P-K) est complété à son extrémité 5' et/ou à son extrémité 3', par un ou plusieurs codons, codant par exemple pour au moins un résidu lysine à l'extrémité N-terminale du polypeptide.

A titre d'exemple, un oligonucléotide préféré selon l'invention est caractérisé en ce qu'il code pour un polypeptide répondant à la formule (P-K), à la formule K-(P-K)<sub>4</sub>, ou à la formule 2K (P-K)<sub>4</sub>.

Selon un mode de réalisation particulier, la composition de cet oligonucléotide correspond à l'une des séquences décrites dans les pages qui suivent et identifiées par les désignations SEQ ID No:1 et SEQ ID No:2.

Les oligonucléotides précédemment décrits constituent des moyens de base pour réaliser des séquences de nucléotides recombinantes capables d'exprimer des protéines ou des polypeptides de réserve de plantes enrichis en lysine.

L'invention a donc aussi pour objet une séquence de nucléotides recombinante comprenant un enchaînement de nucléotides codant pour une protéine de réserve d'une plante, caractérisée en ce qu'elle comprend en outre un oligonucléotide selon l'invention, inséré en un site de l'enchaînement de nucléotides choisi de façon telle que :

- l'expression de la séquence de nucléotides dans une cellule végétale déterminée permet d'obtenir une protéine de réserve modifiée localisée de façon identique ou similaire dans cette cellule à la protéine de réserve normale qui serait exprimée dans la même cellule, dans les mêmes conditions par l'enchaînement de nucléotides normal codant correspondant, et/ou

- la protéine de réserve modifiée codée par la séquence de nucléotides est reconnue immunologiquement par des anticorps produits contre la protéine de réserve normale correspondante.



En particulier, les anticorps précités sont constitués par un sérum polyclonal ou sont obtenus contre des épitopes de la protéine de réserve normale qui sont conservés dans la protéine de réserve modifiée.

Les cellules végétales auxquelles il est fait référence ci-dessus  
5 comprennent toute cellule végétale, quelle que soit son origine tissulaire ou sa nature. En particulier on s'intéresse dans le cadre de l'invention aux cellules des organes de réserve, mais aussi aux cellules de feuilles, tiges, tubercules ...

Par l'expression "protéine de réserve" d'une plante, on entend dans  
10 le cadre de la présente demande une protéine synthétisée durant la maturation des semences et qui est utilisée pendant la phase de germination comme réserve principale de nutrition.

De façon générale, il s'agit d'un polypeptide susceptible d'être synthétisé dans un tissu de réserve quelle qu'en soit la localisation dans la  
15 plante ; les protéines de réserve mises en oeuvre dans le cadre de la présente demande sont en particulier celles produites dans les graines ou les semences des plantes de la famille des céréales, des crucifères ou des légumineuses, et sont par exemple les prolamines, les zéïnes.

Le choix du/des site(s) d'insertion de l'oligonucléotide dans  
20 l'enchaînement codant pour la protéine de réserve de plante est déterminé par le respect des conditions ci-dessus énoncées. Selon le cas, l'insertion peut avoir lieu dans un domaine répétitif (en termes de séquence d'acides aminés) de la protéine ou au niveau d'une extrémité C- ou N-terminale.

La condition donnée ci-dessus selon laquelle l'expression de la  
25 séquence de nucléotides recombinante de l'invention dans une cellule végétale permet d'obtenir une protéine de réserve modifiée, localisée identiquement ou similairement à la protéine de réserve normale qui serait exprimée dans les mêmes conditions dans la même cellule végétale, comporte par exemple pour les  $\gamma$ -zéïnes synthétisées, la possibilité d'être  
30 accumulées dans le réticulum endoplasmique des cellules végétales qui



l'expriment, et en particulier dans les corps protéiques formés à partir de ce réticulum endoplasmique, lorsque la protéine est exprimée dans les cellules de l'endosperme.

Pour obtenir ce résultat au moyen des séquences de nucléotides recombinautes de l'invention, des systèmes d'expression adaptés à l'hôte cellulaire dans lequel est exprimée la séquence de nucléotides sélectionnée, et notamment les éléments de régulation, par exemple les promoteurs, sont choisis pour leur caractère fonctionnel dans le tissu contenant les cellules transformées. Des tests pour effectuer ce choix peuvent être réalisés en s'appuyant sur les différentes constructions décrites dans les exemples.

Pour vérifier que les propriétés immunologiques de la protéine de réserve modifiée exprimée par la séquence de nucléotides de l'invention n'ont pas été modifiées de façon conséquente, on a utilisé par exemple des antisérums tels que l'antisérum  $\alpha$ G2, décrit précisément dans la partie expérimentale qui suit.

Selon un premier mode de réalisation de l'invention, la séquence de nucléotides recombinante est caractérisée en ce qu'elle est obtenue à partir d'un enchaînement codant de nucléotides qui conduit à l'expression d'une protéine de réserve naturellement pauvre en lysine.

De façon générale, cette séquence de nucléotides recombinante code pour une protéine de réserve modifiée dérivée d'une protéine de réserve naturellement produite par une plante utilisable en alimentation animale ou humaine.

Ainsi, les protéines de réserve dont on modifie la teneur en lysine dans le cadre de la présente invention sont avantageusement des protéines de réserve de plantes de la famille des céréales, des légumineuses ou des crucifères. Des protéines de réserve particulièrement importantes sont celles du maïs, en particulier les zéines, et plus



spécialement la  $\gamma$ -zéine du maïs, dont on cherche à augmenter la teneur en lysine.

Une séquence de nucléotides recombinante particulière selon l'invention est caractérisée en ce que l'enchaînement codant de  
5 nucléotides codant pour la  $\gamma$ -zéine de maïs qu'elle contient, répond à la séquence présentée sur la Figure 9.

D'autres séquences de nucléotides recombinantes de l'invention sont caractérisées en ce que l'enchaînement codant de nucléotides qu'elles comprennent, code pour une protéine de réserve d'une plante  
10 choisie parmi les suivantes : le soja, le tournesol, le tabac, le blé, l'avoine, la luzerne, le riz, le colza, le sorgo, Arabidopsis.

Selon un mode de réalisation préféré de l'invention, dans la séquence de nucléotides recombinante comprenant un enchaînement codant pour la  $\gamma$ -zéine de maïs, l'oligonucléotide de l'invention est inséré à  
15 la place de l'enchaînement codant pour le domaine Pro-X naturellement présent dans la séquence d'acides aminés de la  $\gamma$ -zéine du maïs ou à la suite de cet enchaînement. Le domaine Pro-X de la séquence d'acides aminés de la  $\gamma$ -zéine du maïs est constitué par les acides aminés situés entre les positions 70 et 91 de la séquence d'acides aminés représentée à  
20 la Figure 9, correspondant aux nucléotides 265 à 330 de la séquence représentée à la Figure 9.

De façon préférée, dans la séquence de nucléotides de l'invention, l'oligonucléotide à la place ou à la suite du domaine Pro-X présent entre les nucléotides 276 et 357 de la séquence représentée à la Figure 9.

25 Selon un autre mode de réalisation de l'invention, dans la séquence de nucléotides recombinante comprenant un enchaînement codant pour la  $\gamma$ -zéine de maïs, l'oligonucléotide de l'invention est inséré à la suite du domaine Pro-X conservé dans la séquence de la  $\gamma$ -zéine du maïs.





Selon une autre variante de réalisation, dans la séquence de nucléotides recombinante comprenant un enchaînement codant pour la  $\gamma$ -zéine de maïs, l'oligonucléotide de l'invention est inséré dans le domaine Pro-X maintenu dans la séquence de la  $\gamma$ -zéine.

5 Les insertions dont il est question ci-dessus sont réalisables par les techniques disponibles, par exemple par recombinaison des séquences ayant subi une ou plusieurs étapes de digestion enzymatique.

Dans un mode de réalisation particulier de l'invention, une protéine de réserve choisie enrichie en un acide aminé déterminé est exprimée  
10 dans des cellules végétales hétérologues. En d'autres termes une protéine de réserve présente naturellement dans une plante donnée, est exprimée sous forme enrichie en acides aminés, dans une autre plante, ou dans une autre cellule que celle qui la produit naturellement.

Outre l'enchaînement codant pour une protéine de réserve de plante  
15 et l'oligonucléotide de l'invention, les séquences de nucléotides recombinantes de l'invention peuvent également comprendre un promoteur d'expression et, par exemple, un promoteur choisi pour son caractère spécifique pour l'expression dans certaines parties ou dans certains tissus des plantes, ou au contraire un promoteur choisi pour son caractère  
20 constitutif. A titre d'exemple, lorsqu'ils sont spécifiques, les promoteurs peuvent être spécifiques des semences et/ou d'organes ou de tissus déterminés de plantes. Ils peuvent alternativement ou également être spécifiques d'une phase de croissance, par exemple d'un stade déterminé de la germination.

25 A l'inverse l'utilisation de promoteurs constitutifs permet l'expression constante et générale de la protéine de réserve, entraînant une compétition entre l'expression de la protéine de réserve native lorsqu'elle est présente et la protéine de réserve modifiée.

A titre d'exemple, des promoteurs avantageux pour la réalisation de  
30 l'invention sont le promoteur de la  $\gamma$ -zéine du maïs contenu dans la



séquence de 1,7 kb se trouvant en amont de la séquence codant représentée à la Figure 7, le promoteur du virus de la mosaïque du chou-fleur, à savoir le promoteur CaMV35S (EP-B-0131623), le promoteur constitutif du gène d'actine 1 de riz (PCT/US 9100073) ou le promoteur  
5 semence spécifique "High Molecular weight glutenine" du blé (Colot V. et al, 1987, EMBO Journal, vol. 6, p. 3559-3564).

Le cas échéant, ces promoteurs sont complétés par d'autres séquences de régulation, et en particulier des activateurs d'expression.

A titre d'exemple, d'autres promoteurs utilisables pour la réalisation  
10 de l'invention sont par exemple le promoteur du gène codant pour la protéine de réserve 2S des semences d'Arabidopsis thaliana, ou encore les promoteurs de la lectine de haricot ou de la  $\beta$ -phaséoline de haricot.

L'introduction supplémentaire d'activateurs de l'expression dans les séquences de régulation des séquences de nucléotides selon l'invention  
15 permet également d'augmenter le niveau de la transcription primaire de la séquence de nucléotides et, le cas échéant, d'augmenter la quantité de protéines de réserves modifiées produites. Des activateurs sont par exemple des introns de monocotylédones tels que l'intron 1 du gène d'actine de riz.

20 L'invention a également pour objet un vecteur de clonage et/ou d'expression, caractérisé en ce qu'il comprend, en un site non essentiel pour sa répllication, une séquence de nucléotides répondant à l'une des définitions précédemment données. Des vecteurs particuliers intéressants dans le cadre de la réalisation de l'invention sont par exemple les  
25 plasmides pP20 $\gamma$ Z, pH30 $\gamma$ Z ou pH45 $\gamma$ Z. Le plasmide pP20 $\gamma$ Z a été déposé à la CNCM (Paris, FRANCE) le 31 Octobre 1995 sous le numéro I-1640. Le plasmide pH45 $\gamma$ Z a été déposé à la CNCM le 31 Octobre 1995 sous le numéro I-1639.



Entre également dans le cadre de l'invention un polypeptide tel qu'exprimé par une séquence de nucléotides recombinante répondant aux définitions précédentes.

L'expression "polypeptide" n'introduit pas dans le cadre de la présente invention de limitation particulière quant au nombre d'acides aminés formant le polypeptide. Il peut donc s'agir de séquences comportant quelques acides aminés, habituellement désignées par l'expression "peptides", ou encore de séquences beaucoup plus longues telles que celles des protéines.

10 L'invention concerne à cet égard la  $\gamma$ -zéine de maïs modifiée riche en lysine, caractérisée en ce qu'elle est codée par une séquence de nucléotides recombinante décrite précédemment.

Selon un mode de réalisation préféré de l'invention, la  $\gamma$ -zéine de maïs modifiée enrichie en lysine est caractérisée en ce que sa séquence  
15 d'acides aminés est modifiée par au moins un polypeptide répondant à la formule  $(P-K)_n$ , dans laquelle :

- n est un entier supérieur ou égal à 2,
- P représente un résidu d'acide aminé Proline,
- K représente un résidu d'acide aminé Lysine,
- 20 - le signe "-" symbolise une liaison entre les deux résidus d'acides aminés et notamment une liaison de type peptidique, les n unités  $(P-K)$  étant liées par des liaisons notamment de type peptidique.

Selon une variante de réalisation de l'invention le polypeptide intégré dans la séquence d'acides aminés de la  $\gamma$ -zéine répond à la  
25 formule  $K-(P-K)_n$ .

Les polypeptides de l'invention répondant à l'une des formules  $(P-K)_n$ ,  $K-(P-K)_n$  ou à des variantes sont substitués à une séquence naturellement présente dans la  $\gamma$ -zéine de maïs normale ou insérés avec délétion d'un ou plusieurs acides aminés de la séquence d'acides aminés



de la  $\gamma$ -zéine de maïs normale, ou ajoutés dans la séquence d'acides aminés de la  $\gamma$ -zéine normale, le site d'insertion du polypeptide étant choisi de façon telle que :

- lorsque la  $\gamma$ -zéine modifiée riche en lysine est produite dans une cellule hôte, notamment une cellule végétale, elle est localisée de façon identique ou similaire dans cette cellule à la  $\gamma$ -zéine de maïs normale qui serait produite dans les mêmes conditions, dans la même cellule hôte et/ou,

- la  $\gamma$ -zéine de maïs modifiée est reconnue par des anticorps dirigés contre la  $\gamma$ -zéine normale de maïs.

Les protéines P20 $\gamma$ Z représentée à la figure 11 ou H30 $\gamma$ Z ou H45 $\gamma$ Z représentée à la figure 10, sont des exemples préférés de la réalisation de l'invention et représentent des  $\gamma$ -zéines de maïs modifiées enrichies en lysine.

L'invention a également pour objet une cellule hôte recombinante, caractérisée en ce qu'elle comprend une séquence de nucléotides décrite ci-dessus.

Des cellules hôtes intéressantes sont par exemple les cellules de bactéries, telles que celles de E. coli ou d'Agrobacterium tumefaciens. De préférence, dans le cadre de l'invention, et pour l'expression stable de la protéine de réserve modifiée recherchée, on aura recours à des cellules hôtes d'origine végétale.

A titre d'exemple, ces cellules d'origine végétale sont des cellules de semences, de plante, et par exemple, à titre préféré, des cellules d'endosperme de semences de maïs.

La séquence de nucléotides de l'invention est de préférence introduite dans le génome de la cellule hôte de façon stable et dans des conditions telles que la protéine de réserve exprimée enrichie en acides





aminés et en particulier en lysine est localisée comme le serait la protéine correspondante normale dans la même cellule hôte.

Des techniques variées sont disponibles pour transformer les cellules hôtes. A titre d'exemple, et en vue de transformer de façon stable ou de façon transitoire les cellules hôtes, on utilisera les techniques d'électroporation, de bombardement de microprojectiles portant de l'ADN, via canon à particules, via culture d'explants avec Agrobacterium tumefaciens, par pénétration de microfibres.

Outre les cellules d'endosperme de semences de maïs, des cellules de soja, de tournesol, de tabac, de blé, d'avoine, de luzerne, de riz, de colza, de sorgo ou d'Arabidopsis peuvent être mises en oeuvre pour exprimer les séquences de nucléotides de l'invention.

La présente demande concerne aussi les semences produisant un polypeptide tel que décrit ci-dessus et les plantes produisant ce même polypeptide. De préférence, ces plantes sont du maïs.

L'invention a aussi pour objet des semences obtenues à partir des plantes transformées exprimant le polypeptide de l'invention, en d'autres termes le polypeptide de réserve modifié enrichi en acides aminés déterminés.

Selon un mode de réalisation particulièrement intéressant de l'invention, les protéines  $\gamma$ -zéines modifiées enrichies en lysine, sont exprimées chez des mutants opaque 2 de maïs. Ces mutants o2 décrits par Emerson R.A. et al. (1935, Cornell Univ. Agric. Exp. Stn. Mem. 180) et caractérisés par Mertz E.T. et al (1964, Science 145 : 279-280) voient leur contenu en lysine fortement augmenté améliorant ainsi fortement les qualités nutritives du maïs (compensation de son bas niveau en cet acide aminé essentiel). Les maïs classiques ont un contenu en lysine d'environ 0,24% du produit brut (poids du grain total), les maïs opaque 2 avoisinent en revanche 0,5% en lysine. Cependant ils présentent des caractéristiques agronomiques insuffisantes car leur endosperme est beaucoup moins



vitreux et se révèle très friable (phénotype "starchy"). Ceci a pour effet de les rendre extrêmement sensibles aux organismes pathogènes et aux étapes de traitements post récolte. Ce phénotype est en fait dû à la diminution importante de certaines protéines de réserve en particulier les  
5 alpha zéines. En fait Opaque 2 code pour un facteur de transcription nécessaire à l'expression de certains gènes zéines (Schmidt R. J. et al, 1990, Proc. Natl. Acad. Sci. USA 87, 46-50).

Des dérivés opaque 2 n'ayant plus les inconvénients précités ont été développés par amélioration génétique classique, il s'agit des maïs  
10 QPM (Quality Protein Maize). L'analyse génétique récente de ces maïs (Lopes M.A. et al, 1995, Theor. Appl. Genet. 19, 274-281) a montré que seulement 2 ou 3 loci se révèlent clefs dans ces modifications favorables. Des analyses génétiques et biochimiques plus poussées permettent de postuler qu'un des 3 loci responsables est le locus  $\gamma$ -zéine : les génotypes  
15 de maïs qui portent une duplication de ce gène situé dans la région centrométrique du chromosome 7, s'avèrent tous être des modificateurs d'opaque 2 (Lopes M.A. et al, 1995, Mol. Gen. Genet. 19 : 247 : 603-613).

La présente invention permet de préparer des maïs mutants opaque 2 à partir de maïs n'ayant qu'un gène de  $\gamma$ -zéine au niveau du chromosome  
20 7, qui sont complémentés par l'addition d'une séquence recombinante codant pour une  $\gamma$ -zéine de maïs enrichie en lysine. Outre le fait d'acquérir les propriétés de dureté similaire à un maïs opaque 2 non mutant, il a l'avantage d'augmenter significativement la teneur en lysine, dépassant ainsi celle des maïs QPM.

25 La présente invention permet d'obtenir des mutants opaque 2 de maïs modifiés, dans lesquels on a inséré une séquence de nucléotides recombinante codant pour une  $\gamma$ -zéine de maïs enrichie en lysine.



L'invention a aussi pour objet un procédé pour l'obtention de plantes ou de semences exprimant une protéine de réserve modifiée, caractérisé en ce qu'il comprend les étapes de :

- a) transformation d'une cellule végétale, avec une séquence de nucléotides ou un vecteur ci-dessus décrits, dans des conditions permettant l'expression de façon stable et fonctionnelle de la protéine de réserve modifiée codée par la séquence de nucléotides;
- b) régénération de plantes à partir de la cellule de plante transformée de l'étape a), pour obtenir des plantes exprimant la protéine de réserve modifiée,
- c) le cas échéant, obtention des semences à partir des plantes modifiées obtenues à l'étape b).

Selon un mode de réalisation avantageux de l'invention, la plante transformée est le maïs et la protéine de réserve modifiée enrichie est la  $\gamma$ -zéine, enrichie en lysine.

L'invention concerne aussi les plantes obtenues par mise en oeuvre d'un tel procédé.

Afin d'évaluer la teneur en un acide aminé donné des plantes selon l'invention, il est possible d'utiliser un protocole de dosage tel que celui mentionné dans Zarkadas et al, 1995, J. Agri. Food. Chem. Vol. 43: pages 84-93.

D'autres caractéristiques et avantages de l'invention apparaissent dans les exemples et dans les figures qui suivent.

**Figure 1**

- Carte de restriction du plasmide pP20 $\gamma$ Z

**Figure 2**

- Carte de restriction du plasmide pH45 $\gamma$ Z



**Figure 3**

- Représentation schématique des protéines codées par les gènes modifiés et non modifiés de la  $\gamma$ -zéine :  $\gamma$ -zéine de type sauvage ( $\gamma$ Z),  $\gamma$ -zéines riches en lysine (P20 $\gamma$ Z, H30 $\gamma$ Z, H45 $\gamma$ Z et N13 $\gamma$ Z) résultant de l'insertion des oligonucléotides codant pour des séquences riches en lysine. La séquence d'acides aminés des polypeptides insérés est indiquée en utilisant le code à une lettre de la représentation des acides aminés. Les abréviations suivantes sont utilisées :
- 10 Term : terminal ;  
ProX DOMAIN : domaine linker proline-Xaa.

- Figure 4** - Analyse in-vitro des  $\gamma$ -zéines riches en lysine. (A) traduction in-vitro et translocation des transcrits correspondant aux  $\gamma$ -zéines modifiées riches en lysine ; lignes 1, 5, 9 et 13 : produits complets de traduction ; lignes 2, 6, 10 et 14 : produits complets de traduction après translocation dans les microsomes canins (CM) ; lignes 3, 7, 11 et 15 : produits de translocation résistant à l'action de la protéinase K (PK) ; lignes 4, 8, 12 et 16 : totalité des produits de traduction après traitement avec la protéinase K en présence de 0,5% Nonidet P40(NP40). (B) Immunoprécipitation des produits de traduction in-vitro correspondant aux protéines  $\gamma$ -zéine et  $\gamma$ -zéine modifiées riches en lysine, en utilisant l'anti-sérum  $\alpha$ PL. Ligne 1 :  $\gamma$ -zéine ; ligne 2 : P20 $\gamma$ Z ; ligne 3 : H30 $\gamma$ Z ; ligne 4 : H45 $\gamma$ Z et ligne 5 : N13 $\gamma$ Z. (C) Même légende que pour (B) mais en utilisant l'antisérum  $\alpha$ G2. Les marqueurs de poids moléculaire (en kilodaltons) sont indiqués sur la gauche.

**Figure 5** - Activité tissu-spécifique du promoteur de la  $\gamma$ -zéine. Des endospermes de maïs, des embryons et des feuilles de maïs ont été co-





transformés par bombardement de particules en utilisant les constructions représentées dans la figure (dans la partie à droite). Les activités relatives de la luciférase (LUC, colonnes en gris) et de la  $\beta$ -glucuronidase (GUS, colonnes hachurées) ont été exprimées sous la forme d'un multiplicateur  
5 des valeurs obtenues avec des projectiles nus  $\pm$  la déviation standard des différents ratios.

**Figure 6** - Expression des  $\gamma$ -zéines riches en lysine dans l'endosperme sub-aleuronique des cellules. (A) Immunoblot avec un antisérum  $\alpha$ PL, de  
10 protéines extraites d'endospermes transformés par pN13 $\gamma$ Z (ligne 2), pH45 $\gamma$ Z (ligne 3), pH30 $\gamma$ Z (ligne 4) et pP20 $\gamma$ Z (ligne 5). Le témoin (ligne 1) correspond à des endospermes non transformés. Les marqueurs de poids moléculaire (en kilodaltons) sont indiqués à gauche. (B) expression des transcrits H45 $\gamma$ Z et N13 $\gamma$ Z dans des endospermes transformés de façon  
15 transitoire. Les cDNAs obtenus à partir des tissus transformés avec pH45 $\gamma$ Z (ligne 2), pN13 $\gamma$ Z (ligne 3) et le témoin (ligne 1) ont été amplifiés par PCR et analysés en utilisant un oligonucléotide synthétique codant pour une séquence riche en lysine à titre de sonde.

20 **Figure 7** - Accumulation de  $\gamma$ -zéines riches en lysine dans les corps protéiques de l'endosperme. (A) Analyse en immunoblot en utilisant l'antisérum  $\alpha$ PL, des corps protéiques isolés à partir des endospermes transformés avec pP20 $\gamma$ Z (ligne 1), pH30 $\gamma$ Z (ligne 2), pH45 $\gamma$ Z (ligne 3), pN13 $\gamma$ Z (ligne 4) et aucun ADN (ligne 5). (B) Analyse en immunoblot en  
25 utilisant l'antisérum  $\alpha$ PL, des corps protéiques isolés à partir des endospermes transformés avec pP20 $\gamma$ Z, pH30 $\gamma$ Z et pH45 $\gamma$ Z et digérés avec la protéinase K en présence d'un tampon isotonique (Sucr., lignes 1, 3 et 5) ou un tampon hypotonique (H<sub>2</sub>O, lignes 2, 4 et 6). Les marqueurs de poids moléculaire (en kilodaltons) sont indiqués sur la gauche.



**Figure 8** - Co-localisation des protéines P20 $\gamma$ Z avec les  $\alpha$ - et les  $\gamma$ -zéines dans les corps protéiques de l'endosperme du maïs. Une analyse en immunocytochimie a été réalisée sur des sections ultrafines en utilisant des anticorps  $\alpha$ PL (marqués avec des particules d'or de diamètre 15 nm) et des anticorps  $\alpha$ Z et  $\alpha$ G2 (marqués avec des particules d'or de 5 nm). (A) Corps protéiques de l'endosperme transformé avec pP20 $\gamma$ Z, immunomarqués avec l'anticorps  $\alpha$ PL. (B) Immunolocalisation de P20 $\gamma$ Z (marqué avec des particules d'or de 15 nm) et de l' $\alpha$ -zéine (marquée avec les particules d'or de 5 nm) dans les corps protéiques isolés à partir des endospermes transformés avec pP20 $\gamma$ Z. (C) et (D) Immunolocalisation de P20 $\gamma$ Z (marqué avec les particules d'or de 15 nm) et des  $\gamma$ -zéines (marquées avec les particules d'or de 5 nm) dans des corps protéiques isolés à partir des endospermes transformés avec pP20 $\gamma$ Z. Les flèches indiquent les sections tangentiellles des corps protéiques.

**Figure 9** - Séquence codante de l'ADNc la  $\gamma$ -zéine de maïs et la séquence d'acides aminés correspondante.

20

**Figure 10** - Séquence codante de l'ADNc de la H45 $\gamma$ Z zéine de maïs et la séquence d'acides aminés correspondante.

La séquence riche en lysine (28 acides aminés) est introduite entre les résidus d'acides aminés 92 et 119 de la séquence représentée à la Figure 10.

25

**Figure 11** - Séquence codante de l'ADNc de la P20 $\gamma$ Z zéine de maïs et la séquence d'acides aminés correspondante.



La séquence riche en lysine (14 acides aminés) est introduite entre les résidus d'acides aminés 92 et 119 de la séquence représentée à la Figure 11.

- 5 **Figure 12** - Cartes de restriction des plasmides pBin 19P20 $\gamma$ Z et pBin 19H30 $\gamma$ Z.

**Figure 13** - Endospermes des plants de maïs transgéniques accumulant de la  $\gamma$ -zéine enrichie en lysine

- 10 A et B: SDS-page et Immunoblot utilisant l'antisérum  $\alpha$ PL

A) 10  $\mu$ g de protéine par piste (transformants avec le construit H45 $\gamma$ Z)  
piste C: extrait protéique des endospermes d'hybrides B73xA188 (contrôle)  
piste 1: A1

- 15 piste 2: B1  
piste 3: B2  
piste 4: C1  
piste 5: D1  
piste 6: D2

20

B) 1  $\mu$ g de protéine par piste (transformants avec le construit P20 $\gamma$ Z)  
piste C: contrôle

- piste 1: A1  
piste 2: A2  
25 piste 3: B1  
piste 4: C1  
piste 5: D1  
piste 6: E1



C) SDS-page et coloration à l'argent (3 transformants P20 $\gamma$ Z et 3 transformants H45 $\gamma$ Z)

piste 1: A2	}	10 µg de protéine par piste
piste 2: B1		
piste 3: C1		
piste 4: B1		
piste 5: A1		
piste 6: D2		
piste 7: B1	}	40 µg de protéine par piste
piste 8: A1		
piste 9: D2		

- 5 **Figure 14** - Contenu en  $\gamma$ -zéine enrichie en lysine par grain (transformants 45 $\gamma$ Z B1 et C1)

A: coloration à l'argent; 10  $\mu$ g de protéine par piste

B: immunoblot utilisant l'antisérum  $\alpha$ PL; 1  $\mu$ g de protéine par piste

- 10 Pistes 1 à 5: extraits protéiques de différents endospermes du transformant 45 $\gamma$ Z B1.

Pistes 6 à 10: extraits protéiques de différents endospermes du transformant 45 $\gamma$ Z C1.

- 15 **Figure 15** :

A) Contenu en  $\gamma$ -zéine enrichie en lysine de 10 grains (transformant 45 $\gamma$ Z C1) utilisant l'antisérum  $\alpha$ PL 1  $\mu$ g de protéine par piste

Pistes 1 à 10: extrait d'endospermes de 10 descendants





B) Immunoblot des extraits protéiques des endospermes 1 à 5 présents en A) et marqués avec l'antisérum  $\alpha$ G2; 2  $\mu$ g de protéine par piste.

### EXEMPLES

5

**A) Préparation de  $\gamma$ -zéines modifiées enrichies en lysine et expression de ces protéines modifiées suivie de l'accumulation dans les corps protéiques des cellules de l'endosperme de maïs**

10 La  $\gamma$ -zéine est une protéine de réserve du maïs, ayant un poids moléculaire de 28 kD, riche en soufre, qui est accumulée dans les cellules de l'endosperme avec les  $\alpha$ -et  $\beta$ -zéines, dans les corps protéiques dérivés du réticulum endoplasmique granulaire (ER) (Ludevid et al., 1984, Plant Mol. Biol. 3, 227-234 ; Lending et al., 1984, Plant Cell 1, 1011-1023). La

15 séquence d'acides aminés déduite de la séquence nucléotidique de l'ADNc (Prat et al., 1985, Nucl. Acids Res. 13, 1493-1504) et des clones génomiques (Boronat et al., 1986, Plant Sci. 47, 95-102) montre que la  $\gamma$ -zéine n'a pas d'homologie avec les polypeptides de type  $\alpha$ -zéine. Bien que la  $\gamma$ -zéine soit codée par 1 à 2 gènes par génome haploïde (Boronat et al.,

20 1986, Plant Sci. 47, 95-102) elle représente 10-15% de la totalité des protéines des endospermes de maïs. L'expression du gène de la  $\gamma$ -zéine dans des systèmes hétérologues tels que les oocytes de *Xenopus* (Torrent et al., 1994, Planta 192, 512-518) et dans *Arabidopsis thaliana* (Geli et al., 1994, Plant Cell 6, 1911-1922), indique que les polypeptides de  $\gamma$ -zéine

25 restent stables et sont capables en tant que tels de former des corps protéiques dérivés du réticulum endoplasmique, à l'intérieur des cellules. De plus, des analyses par délétion de différents domaines structuraux de la  $\gamma$ -zéine ont montré que la séquence N-terminale incluant le domaine répété riche en proline, était responsable de la rétention de la  $\gamma$ -zéine dans



le réticulum endoplasmique et que le domaine C-terminal riche en cystéine était responsable de la formation des corps protéiques. Le domaine Pro-X ne semblait pas en relation avec la stabilité de la protéine ni avec sa localisation ciblée (Geli et al., 1994, Plant Cell 6, 1911-1922).

## 5 Matériel et Méthodes

### **Matériel végétal**

Après une stérilisation en surface (1) des graines de stade 17 DAP  
10 ("days after pollination") du maïs de sélection W64A ont été disséquées à la main et la couche de péricarpe et d'aleurone a été séparée des endospermes. Des sections tangentiellles ont été réalisées de façon à exposer une grande partie de la surface sub-aleuronique. Si nécessaire, des embryons ont été isolés et des feuilles de plantes âgées de 7 jours ont  
15 été disséquées pour retirer le tissu épidermique. Après dissection, les échantillons ont été déposés dans des boîtes de Petri sur des papiers filtres humidifiés avec un milieu MS (Murashige et Skoog, 1962, Physiol. Plant. 15, 473-497).

## 20 **Constructions plasmidiques**

Un premier ensemble de plasmides, pKSG2, pHbP2, pPbP4 et pNaN1 a été obtenu pour permettre l'introduction de sites de restriction dans le gène codant pour la  $\gamma$ -zéine. pKSG2 et pHbP2 ont été construits  
25 selon la description décrite dans la publication de Torrent M. et al. (Planta (1994) 192 : 512-518). Le plasmide pKSG2 contient la séquence codante de la  $\gamma$ -zéine.



Le plasmide pHbP2 est obtenu à partir de pKSG2 et contient une séquence codant pour une  $\gamma$ -zéine mutée dans laquelle le domaine Pro-X de la protéine a été délété.

Le plasmide pPbP4 a été obtenu à la suite de deux étapes de clonage : (i) le fragment de restriction Sall-PvuII de 350 pb de pKSG2 a été cloné dans un plasmide Bluescript (pBSKS, Stratagene, La Jolla, California USA) restreint avec Sall et EcoRV (pKSC4) et (ii) le fragment de restriction PvuII-XbaI de 600 bp de pKSG2 a été cloné dans les sites de restriction SmaI-XbaI de pKSC4. La nouvelle construction pPbP4 contient un site de restriction utile EcoRI juste avant le domaine P-X de la séquence codante de la  $\gamma$ -zéine.

Le plasmide pNaN1 a également été obtenu à l'aide de deux étapes de clonage : (i) le fragment NaeI-XbaI de 250 bp de pKSG2 a été cloné dans le plasmide pBSKS restreint avec EcoRV-XbaI (pKSC8) et (ii) le fragment de restriction NaeI-HindIII de 700 bp (à extrémités franches) de pKSG2 a été cloné dans le site de restriction HindIII de pKSC8. La nouvelle construction, pNaN1, contient des sites de restriction ClaI et HindIII à une position située à 15 nucléotides avant le codon stop de la  $\gamma$ -zéine.

Deux oligonucléotides synthétiques répondant aux séquences suivantes : SEQ ID N°1 :

5'CGATGAATTCAAACCAAAGCCAAAGCCGAAGCCAAAAGAATTCA3', et sa séquence inverse appelée SEQ ID N°2, dont la séquence est la suivante :

5'AGCTTGAATTCTTTTGGCTTCGGCTTTGGCTTTGGTTTGAATTCAT3' codant pour des séquences riches en lysine désignées K(P-K)<sub>4</sub>, ont été hybridées, digérées avec EcoRI et clonées dans un site EcoRI de pHbP2 et pPbP4. Trois clones ont été sélectionnés : pPo2 et pHo3 contenant la séquence codante pour K(P-K)<sub>4</sub> et pHo4 comprenant la forme tronquée de



la séquence codant pour le  $\gamma$ -zéine contenant un tandem  $2K(P-K)_4$  (sous forme d'une séquence  $K(P-K)_4$  EF  $K(P-K)_4$ ) de la séquence codant riche en lysine. Les mêmes oligonucléotides hybridés ont été digérés avec des enzymes *Clal-HindIII* et clonés dans le plasmide *pNaN1* restreint à l'aide  
5 des mêmes enzymes. Le clone sélectionné, *pNo1*, contenait la séquence codant pour la séquence riche en lysine  $K(P-K)_4$  à l'extrémité N-terminale de la  $\gamma$ -zéine modifiée correspondante.

Pour la transformation transitoire de l'endosperme, les séquences codant pour la  $\gamma$ -zéine modifiée de *pPo2* et *pHo3* ont été insérées sous la  
10 forme de fragments *HincII-NheI* dans des sites *SmaI-XbaI* de *pDH51* (Pietrzah et al., 1986, Nucl. Acid Res. 14, 5857-5868) contenant le promoteur 35S du virus de la Mosaïque du chou-fleur (CaMV). Le promoteur *pP20 $\gamma$ Z* obtenu par l'insertion décrite ci-dessus des fragments *HincII-NheI* dans les plasmide *pDH51* contient la séquence codante de la  
15  $\gamma$ -zéine enrichie en lysine (Figure 8) ainsi que les signaux de la séquence 35S du virus CaMV pour la formation de l'extrémité 3' et la polyadénilation. La séquence codante chimère *P20 $\gamma$ Z* a été construite à partir de la région codante de la  $\gamma$ -zéine contenu dans le plasmide *pKSG2* après différentes étapes de clonage. Le promoteur de 1,7 kb de la  $\gamma$ -zéine (Reina et al.  
20 1990, Nucl. Acids Res. 18, 6426) a été inséré dans les extrémités franches d'un fragment *HindIII-PvuI* dans *pHo4* et *pNo1* restreints avec *Xho1* et obtenus avec des extrémités franches. Les constructions *pH45 $\gamma$ Z* et *pN13 $\gamma$ Z* ont respectivement été obtenues.

Les nouvelles constructions, respectivement appelées *pP20 $\gamma$ Z*,  
25 *pH30 $\gamma$ Z*, *pH45 $\gamma$ Z* et *pN13 $\gamma$ Z* ont été utilisées dans des expériences de bombardement biolistique.

Pour étudier la spécificité de différents promoteurs vis-à-vis des tissus végétaux, deux constructions, *p1.7 $\gamma$ ZGUS* et *pCaMV35SLUC* ont été utilisé s. *p1.7 $\gamma$ ZGUS* a été obtenue par insertion du promoteur de la  $\gamma$ -





zéine de 1,7 kb (HindIII-PvuI) dans un plasmide dérivé de pPuC18 contenant le gène GUS et les signaux NOS de polyadénilation en 3' de pBI 101.1 (Jefferson et al., 1987, Embo. J. 6, 3901-3907). pCaMV35SLUC a été obtenu par insertion du gène codant pour la luciférase (LUC) de  
5 pAHC18 (Bruce et al., 1989, P. H. 86, 9692-9696) dans le polylinker de pDH51 (Pietrzak et al., 1986, Nucl. Acids Res. 18, 6426).

### Analyse in-vitro

10 Les plasmides dérivés de pBSKS contenant les séquences codantes de la  $\gamma$ -zéine (pKSG2) et de la  $\gamma$ -zéine riche en lysine (pPo2, pHo3, pHo4 et pNo1) ont été transcrits in-vitro suivant les protocoles standards (Sambrook et al., 1989, Molecular Cloning: A laboratory manual, 2nd. ed., Cold Spring Harbor Laboratory Ed., Cold Spring Harbor, New  
15 York). Une traduction in-vitro et une translocation des transcrits synthétiques ont été réalisées selon la technique de Torrent et al., (1994, Panta 192, 512-518), à l'exception des microsomes canins (CM) utilisés qui provenaient de Promega (Madison, Wis., USA). L'immunoprécipitation des produits traduits a été effectuée essentiellement selon la méthode de  
20 Borgese et Gaetani (1980) en utilisant un sérum de lapin anti- $\gamma$ -zéine $\alpha$ -G2, (Ludevid et al., 1985, Plant Sci. 41, 41-48) et un antisérum  $\alpha$ PL.  $\alpha$ PL est un antisérum polyclonal de lapin obtenu contre le peptide synthétique EFK(P-K)<sub>8</sub>EF. Ce peptide a été synthétisé en utilisant la technique de synthèse en phase solide décrite par Celma et al., 1992.

25

### Bombardement de microprojectile

L'ADN plasmidique a été absorbé sur des particules d'or (1,0  $\mu$ m, Bio-Rad, Lab., Richmond, CA, USA) selon un protocole décrit par Kikkert



(Plant Cell, 33:221-226, 1993). Toutes les cibles ont été bombardées deux fois, en utilisant le dispositif BioRad Biolistic PDS-1000/He. Les cibles ont été positionnées 8 cm derrière un écran stoppant les macroporteurs, qui étaient positionnés 1 cm au dessous du disque de rupture 900 PSI. Après  
5 le bombardement, les échantillons ont été incubés pendant 24 heures à 26°C dans le noir. Des témoins étaient constitués par les cibles bombardées avec des microprojectiles dépourvus d'ADN.

### Tests enzymatiques

10

Les tissus bombardés avec les plasmides p1.7 $\gamma$ ZGUS et pCaMV35SLUC ont été homogénéisés sur de la glace dans un tampon contenant 25mM Tris, à pH 7,8, 2mM DTT, 10% de glycérol et 1% de Triton X-100. Après centrifugation à 12.000xg pendant 5 minutes, les  
15 surnageants ont été décantés et la protéine soluble totale dans les extraits a été quantifiée en utilisant le test Bradford (Bio-Rad). L'activité GUS a été testée par analyse fluorimétrique selon la description de Jefferson (1987) en utilisant la 4-méthyl-ombélliféryl- $\beta$ -D-glucuronide (MUG) en tant que substrat. L'activité LUC a été déterminée en utilisant le système de test  
20 Luciférase (Luciferase Assay System Kit) diffusé par Promega, selon les instructions du fabricant.

### Extraction des protéines de réserve et analyse sur gel des protéines

25

Des endospermes transformés avec pP20 $\gamma$ Z, pH20 $\gamma$ Z, pH45 $\gamma$ Z et pN13 $\gamma$ Z ont été réduits à l'état de farine et les  $\alpha$ -zéines ont été extraites à l'aide de trois séries de solvants contenant 70% d'éthanol. La farine résiduelle a été séchée à l'air, et les protéines totales ont été extraites avec un tampon contenant 0,25 M Tris-HCl, pH 6,8, 4% sodium dodecyl



sulfate (SDS) et 5% 2-mercaptoethanol, pendant 1 heure à température ambiante. Les extraits de protéines ont été analysés par SDS-PAGE et immunoblot selon la description de Ludevid et al., 1985. Des feuilles de nitrocellulose ont été intubées avec de l'antisérum  $\alpha$ PL (dilution 1:500) et de la peroxidase de Raifort conjuguée avec un anticorps secondaire (ECL Western Blotting System, Amersham, Buckinghamshire, UK) a été utilisée pour la détection de la protéine.

### Analyse de l'expression de l'ARN

10

L'ARN total a été extrait selon la description de Logemas et al., 1987. L'ADN complémentaire (ADNc) a été préparé en utilisant la transcriptase réverse et l'oligo dT de Gibco BRL (Gaithersburg, MD, USA) selon les instructions du fabricant et cet ARN a été amplifié par une réaction de PCR. Des oligonucléotides amorces utilisées pour la PCR étaient des séquences de 20-mers correspondant aux extrémités de 5' et 3' du gène de structure de la  $\gamma$ -zéine. Des protocoles standards ont été utilisés pour la préparation de sondes marquées au  $^{32}$ P, et pour l'analyse sur gel de l'ADN (Sambrook et al. Molecular Cloning: A laboratory manual, 2nd. ed., Cold Spring Harbor Laboratory Ed., Cold Spring Harbor, New York) en utilisant un oligonucléotide synthétique codant pour une séquence riche en lysine (voir ci-dessus) en tant que sonde.

15  
20

### Isolement des corps protéiques et traitement par la protéase

25

Ces protocoles ont été décrits antérieurement (Torrent et al., Planta, 180: 90-95, 1989).

### Microscopie électronique



Les corps protéiques des endospermes de type sauvage et des endospermes transformés par pP20 $\gamma$ Z ont été fixés avec 2,5% de paraformaldéhyde dans 20 mM de tampon phosphate à pH7,2, pendant 1  
5 heure à température ambiante, et transformés selon la description de Geli et al., 1994, Plant Cell 6, 1911-1922) toutefois en utilisant un antisérum  $\alpha$ -PL et un colloïde d'or et de protéine A ayant un diamètre de 15 nm. Pour le marquage double, des sections ultra fines ont été d'abord incubées avec  $\alpha$ -PL et le colloïde d'or et de protéine A (15 nm de diamètre) a été utilisé  
10 pour la détection de l'anticorps. Après lavage, des sections ont été incubées avec 0,15 mg/mL de protéines A pendant 20 minutes pour saturer les immunoglobulines et finalement les grilles ont été incubées avec des séra  $\alpha$ -G2 ou  $\alpha$ -Z1 et le colloïde or/protéine A (diamètre 5 nm) a été utilisé pour la détection de l'anticorps.  $\alpha$ -Z1 est un antisérum polyclonal  
15 de lapin dirigé contre l' $\alpha$ -zéine obtenu selon la description de Ludevid et al., 1985, Plant Sci. 41, 41-48).

## Résultats

### 20 Construction des $\gamma$ -zéines riches en lysine

Les inventeurs avaient démontré l'importance du domaine répété riche en proline ("proline-rich repeat") et du domaine C-terminal riche en cystéine pour la rétention dans le réticulum endoplasmique de la  $\gamma$ -zéine et  
25 la formation des corps protéiques contenant ces protéines dans les cellules de feuilles de Arabidopsis (Geli et al., 1994, Plant Cell 6, 1911-1922). Sur la base de ces résultats préalables, la possibilité d'insérer des séquences riches en lysine dans différents domaines de la  $\gamma$ -zéine, de façon à créer une  $\gamma$ -zéine modifiée correctement ciblée et accumulée dans





les cellules de l'endosperme a été investiguée afin d'améliorer l s qualités nutritives du maïs.

Les inventeurs ont à présent construit des gènes modifiés de la  $\gamma$ -zéine par introduction d'oligonucléotides synthétiques codant pour des séquences riches en lysine dans différents sites de la séquence codante de la  $\gamma$ -zéine. Des constructions de  $\gamma$ -zéine modifiée ont été créées de façon à éviter le placement des séquences codantes riches en lysine dans les domaines constitués par le domaine répété en tandem et le domaine riche en cystéine. Des modifications de la séquence codante de la  $\gamma$ -zéine ont été effectuées dans la séquence correspondant au domaine Pro-X. De plus pour minimiser une altération du repliement de la protéine, les séquences riches en lysine  $(P-K)_n$  ont été définies pour imiter la séquence du domaine Pro-X. Comme on le voit sur la figure 3, dans la protéine P20 $\gamma$ Z, une séquence  $K(P-K)_4$  a été introduite après la région Pro-X et dans la protéine H30 $\gamma$ Z et dans la protéine H45 $\gamma$ Z, des séquences d'acides aminés comprenant  $K(P-K)_4$  et  $2K(P-K)_4$  respectivement remplacent le domaine Pro-X de la  $\gamma$ -zéine ( $\gamma$ Z, Fig. 3). Pour explorer si l'extrémité C-terminale était un site permissif pour l'introduction des séquences riches en lysine, une protéine supplémentaire N13 $\gamma$ Z, a été créée par l'insertion d'une séquence contenant  $K(P-K)_4$  cinq acides aminés en amont de l'extrémité C-terminale (Fig.3).

#### Activité du promoteur de la $\gamma$ -zéine dans l'endosperme de maïs transformé

Pour déterminer si des  $\gamma$ -zéines riches en lysine pouvaient être exprimées dans les cellules de l'endosperme, on a tout d'abord recherché un promoteur efficace et un système de transformation. Un promoteur de la  $\gamma$ -zéine, contenant une séquence en amont de 1,7 Kb (Reina et al., 1990, Nucleic Acid Research, vol. 18, p. 6426) et le promoteur CaMV contenant



625 bp de la séquence en amont de la protéine 35S du virus de la mosaïque du chou-fleur CaMV ont été testés. Jusqu'à présent aucune information n'était disponible sur l'analyse fonctionnelle des fusions de gènes avec le promoteur de la  $\gamma$ -zéine dans les plantes monocotylédones transgéniques. Pour analyser l'activité et la spécificité tissulaire du promoteur de la  $\gamma$ -zéine, deux gènes chimères ont été construits (voir Fig. 5). l'expression transitoire par bombardement biolistique (Klein et al., 1988 PNAS 85:4305) a été utilisée comme procédure de transformation du maïs pour l'analyse du promoteur aussi bien que pour des expériences d'expression des  $\gamma$ -zéines riches en lysine. Des endospermes de maïs de stade 17 DAP (17 jours après la polynisation) (le péricarpe et les cellules de la couche aleuroniques étaient enlevés), des embryons (17 DAP) et des jeunes feuilles (âgées de 10 jours) ont été bombardés avec des projectiles d'or recouverts avec de l'ADN plasmidique contenant les deux constructions. La figure 5 montre les activités de  $\beta$ -glucuronidase (GUS) et de luciférase (LUC) présentes dans les trois tissus testés : l'endosperme, l'embryon et la feuille, par rapport à l'expérience témoin. On doit noter que les résultats correspondent à la moyenne d'au moins 3 expériences indépendantes réalisées. Toute l'activité GUS sous le contrôle du promoteur de la  $\gamma$ -zéine était restreinte à l'endosperme, puisque aucune expression GUS n'a été détectée dans l'embryon et dans les feuilles. De plus, les endospermes bombardés ont été colorés histochimiquement pour déterminer le nombre de clusters cellulaires exprimant la protéine GUS. Le profil de coloration corroborait les résultats précédents, GUS était fortement exprimé dans les endospermes (le nombre moyen de clusters cellulaires colorés pour GUS par endosperme était de 150) et des tâches bleues n'ont pas été détectées dans l'embryon et dans les feuilles. Au contraire, le promoteur CaMV 35S conférait une activité LUC dans tous les tissus testés (Fig. 5), mais il existait cependant des différences



quantitatives entre l'activité relative de l'enzyme dans les feuilles et les  
embryons par rapport à l'endosperme. Ces différences pouvaient être  
attribuées à une variabilité intrinsèque de l'activité constitutive du  
promoteur CaMV parmi les différents tissus du maïs ou à une pénétration  
5 faible des particules recouvertes par l'ADN dans les cellules mésophyles  
contenant un système vacuolaire important. Il existe dans l'état de la  
technique des preuves selon lesquelles le promoteur CaMV aurait  
habituellement une faible activité dans des cellules de plantes  
monocotylédones (Fromm et al., 1985, Proc. Natl. Acad. Sci. USA 82,  
10 5824-5828) ; cependant les inventeurs ont montré ici une forte activité du  
promoteur CaMV dans des cellules d'endospermes. Ceci a permis de  
conclure que l'activité du promoteur de la  $\gamma$ -zéine et du promoteur CaMV  
35S était très forte dans les endospermes de maïs et ainsi que les deux  
promoteurs pouvaient s'avérer utiles pour contrôler l'expression de la  
15 protéine dans ce tissu.

Pour déterminer si les protéines mutantes codées par les  
constructions étaient compétentes avec la fonction de translocation  
membranaire, des expériences de transcription-traduction in-vitro ont été  
effectuées en présence de microsomes pancréatiques de chien. Les  
20 transcrits synthétiques de chaque construction ont été traduits et la  
translocation à travers les membranes de microsomes testée par  
recherche de la protection vis-à-vis de la digestion par la protéinase K.  
Ces résultats rapportés à la figure 4A indiquent que les poids moléculaires  
apparents des polypeptides synthétisés in-vitro reflètent les mutations  
25 introduites (Fig. 4A, lignes 1, 5, 9 et 13). En présence des microsomes et  
de la protéinase K, des peptides de faible poids moléculaire n'ont pas été  
observés, indiquant que les chaînes polypeptidiques complètes des  $\gamma$ -  
zéines modifiées étaient transportées à travers les membranes  
microsomales (Fig. 4A, lignes 3, 7, 11 et 15). En comparant le résultat de  
30 la translocation des quatre  $\gamma$ -zéine modifiées, on a observé que la protéine



H45 $\gamma$ Z (Fig. 4A, ligne 11), qui contenait l'insertion de 10 acides aminés de type lysine, a subi une translocation inférieure aux autres protéines. Il semble que les résidus chargés négativement pourraient interférer dans une certaine proportion, sur l'efficacité de la translocation. L'anticorps polyclonal  $\alpha$ G2, dirigé contre la  $\gamma$ -zéine (Ludevid et al., 1985, Plant Sci. 41, 41-48) ne pouvant être utilisé pour distinguer entre la  $\gamma$ -zéine de type sauvage et les  $\gamma$ -zéines modifiées, un anticorps  $\alpha$ PL dirigé contre un peptide synthétique contenant la séquence d'acides aminés riche en lysine a été préparé. On a ensuite testé si des protéines modifiées synthétisées in-vitro étaient reconnues à la fois par les anticorps  $\alpha$ G2 et les anticorps  $\alpha$ PL. Cette expérience est illustrée à la figure 4B et la figure 4C où les transcrits synthétiques de la  $\gamma$ -zéine, P20 $\gamma$ Z, H30 $\gamma$ Z, H45 $\gamma$ Z et N13 $\gamma$ Z étaient traduits in-vitro et dans lesquels les produits de traduction étaient immunoprécipités avec  $\alpha$ PL (Fig. 4B) et avec  $\alpha$ G2 (Fig. 4C). Ces résultats indiquent que les  $\gamma$ -zéines riches en lysine étaient reconnues par les deux anticorps (voir Fig. 4B et C, lignes 2 à 5) et que la  $\gamma$ -zéine était seulement reconnue par le sérum  $\alpha$ G2 (Fig. 4B et C, ligne 1). Ainsi la spécificité des anticorps  $\alpha$ PL pour les protéines modifiées a permis de distinguer la  $\gamma$ -zéine riche en lysine de la  $\gamma$ -zéine endogène lorsque les gènes modifiés étaient exprimés dans des cellules d'endosperme. Prises ensemble ces expériences ont montré que la présence de séquences riches en lysine ne perturbait pas la fonction de translocation à travers la membrane ou le comportement immunologique de la  $\gamma$ -zéine.

#### Analyse de l'expression des $\gamma$ -zéines enrichies en lysine dans les endospermes de maïs

Pour explorer si la  $\gamma$ -zéine modifiée riche en lysine était exprimée et accumulée dans les cellules de l'endosperme, des graines de stade 17





DAP ont été bombardées avec l'ADN contenant les séquences codant pour la protéine des quatre constructions (Fig. 3) sous le contrôle du promoteur CaMV (P20 $\gamma$ Z et H20 $\gamma$ Z) et du promoteur de la  $\gamma$ -zéine (H45 $\gamma$ Z et N13 $\gamma$ Z). Des constructions comprenant des promoteurs anti-sens ou dépourvues de promoteurs ont été utilisées comme témoins. Après 24 heures de transformation des endospermes, les protéines totales ont été extraites et l'expression des  $\gamma$ -zéines modifiées a été testée par immunoblotting en utilisant l'anticorps  $\alpha$ PL. La figure 6A montre que les gènes chimères de  $\gamma$ -zéine contenant l'insertion (Pro-Lys)<sub>n</sub> après le domaine Pro-X (P20 $\gamma$ Z) ou le remplaçant (H45 $\gamma$ Z et H30 $\gamma$ Z) étaient fortement exprimés et les produits de traduction accumulés efficacement dans les cellules de l'endosperme (Fig. 6A, lignes 3, 4 et 5). Pour chaque ligne, les extraits de protéine correspondaient à 1/3 environ d'un endosperme bombardé, ce qui a permis d'estimer que la quantité de protéines modifiées P20 $\gamma$ Z, H30 $\gamma$ Z et H45 $\gamma$ Z par endosperme atteignait le niveau du nanogramme. De plus, aucune différence quantitative entre les niveaux d'expression des gènes chimères sous le contrôle du promoteur CaMV et du promoteur de la  $\gamma$ -zéine n'a été observée, confirmant les résultats ci-dessus décrits obtenus avec les protéines marqueurs GUS et LUC (Fig. 5). On a noté que l'anticorps  $\alpha$ PL reconnaissait une protéine d'environ 30 kD présente dans les extraits de protéine totale, même dans les endospermes non transformés (voir la bande faible présente dans les quatre lignes de la figure 6A). Une procédure d'extraction séquentielle de protéine a permis de constater que cette protéine n'était pas une protéine de réserve soluble dans un milieu aqueux.

Comme le montre la figure 6A (ligne 2) aucune trace de la protéine N13 $\gamma$ Z n'a été détectée, indiquant que le gène chimère correspondant n'était pas exprimé dans les cellules de l'endosperme ou que la protéine N13 $\gamma$ Z était dégradée. Des ARNs des endospermes transformés avec les



ADNs codant pour les protéines H45 $\gamma$ Z et N13 $\gamma$ Z ainsi que les ARNs des endospermes non transformés ont été analysés. A partir des ARNs totaux, les ADNc ont été préparés et amplifiés par PCR en utilisant des amorces spécifiques. La Fig. 6B montre l'analyse en Southern blot de trois échantillons d'ADNc hybridés avec un oligonucléotide codant pour une séquence K(Pro-Lys)<sub>4</sub> utilisée comme sonde. Les résultats indiquent que le gène N13 $\gamma$ Z était correctement exprimé (Fig. 6B, ligne 3). La présence de bandes dans les échantillons H45 $\gamma$ Z et N13 $\gamma$ Z, mais pas dans les endospermes non transformés, a suggéré que la protéine N13 $\gamma$ Z était dégradée pendant les 24 heures de l'incubation. A partir de ces observations, les inventeurs ont conclu que le site d'insertion des séquences riches en lysine était critique pour la stabilité de la  $\gamma$ -zéine modifiée.

15 La  $\gamma$ -zéine enrichie en lysine est accumulée dans le corps protéiques

En dehors de la teneur en lysine des séquences Pro-X, les protéines P20 $\gamma$ Z, H30 $\gamma$ Z et H45 $\gamma$ Z avaient des caractéristiques communes avec la  $\gamma$ -zéine de type sauvage : le peptide signal, le domaine répété en tandem N-terminal et la région riche en cystéine C-terminale étaient apparentées. Il a semblé important de déterminer si ces domaines restaient totalement fonctionnels préservant le ciblage et la formation des corps protéiques ou si les séquences riches en lysine créaient un environnement spécial dans lequel ces propriétés pouvaient être perturbées. Pour tester ceci, on a investigué si les  $\gamma$ -zéines modifiées étaient capables de s'accumuler dans les corps protéiques. Un fractionnement sub-cellulaire a été réalisé avec les endospermes transformés. Les homogénats d'endospermes bombardés ont été chargés sur des gradients de sucrose discontinus (20 %, 50 % et 70 % de sucrose)



et toutes les fractions collectées ont été analysées par immunoblotting. P20 $\gamma$ Z, H30 $\gamma$ Z et H45 $\gamma$ Z sédimentés sur la fraction des corps protéiques et aucune quantité significative de ces protéines n'a été détectée, soit dans le surnageant, soit dans la fraction microsomale (Fig. 7A, lignes 2, 2 et 3).

5 Bien que dans les expériences faites préalablement in-vitro (Fig. 4A) aient permis de constater que les  $\gamma$ -zéines modifiées nouvellement synthétisées subissaient une translocation dans les microsomes canins, on a testé ici si les protéines modifiées exprimées in-vivo dans des cellules d'endosperme, subissaient une translocation dans la membrane du réticulum

10 endoplasmique et restaient à l'intérieur des corps protéiques dérivés du réticulum endoplasmique. Pour cette raison, des corps protéiques isolés ont été soumis à une digestion par la protéinase K dans des tampons isotoniques (contenant 20 % de sucrose) ou après un choc osmotique dans l'eau (Fig. 7B). Les protéines protégées contre la dégradation

15 protéolytique par des enzymes peuvent être entourées par une membrane et un traitement par des solutions détergeantes ou hypotoniques a résulté dans la digestion des protéines (Walter et Blobel, 1983, Method Enzymol. 96, 84-93). Une comparaison des intensités de bandes, après digestion par la protéinase K dans des milieux comprenant du sucrose ou de l'eau, a

20 permis de révéler que les protéines P20 $\gamma$ Z, H30 $\gamma$ Z et H45 $\gamma$ Z étaient protégées de la digestion dans des tampons isotoniques (lignes 1, 3 et 5) mais étaient partiellement digérées dans l'eau (lignes 2, 4 et 6).

L'expression des gènes modifiés de la  $\gamma$ -zéine dans les cellules de la couche du sub-aleurone de l'endosperme par bombardement biolistique

25 a permis d'observer que des  $\gamma$ -zéines riches en lysines étaient largement accumulées à l'exception du cas où les séquences riches en lysine étaient positionnées 5 résidus en amont de l'extrémité C-terminale du polypeptide de la  $\gamma$ -zéine. A partir de cette expression et d'études immunocytochimiques sur des corps protéiques isolés, les inventeurs ont



démontré que des  $\gamma$ -zéines riches en lysine sont correctement accumulées dans ces organites et sont co-localisées avec les protéines endogènes  $\gamma$ -zéine et  $\alpha$ -zéine.

Des corps protéiques isolés à partir des endospermes P20 $\gamma$ Z ont été  
5 examinés par marquage de type immunogold et microscopie électronique. Sur des sections ultra fines incubées avec l'anticorps  $\alpha$ P-L (Fig. 8A), le marquage à l'or a été détecté à l'intérieur des corps protéiques, indiquant que la protéine P20 $\gamma$ Z riche en lysine était accumulée à l'intérieur de ces organites. Dans des sections incubées seulement avec l'anticorps  $\alpha$ P-L,  
10 l'immunomarquage avait lieu uniquement sur quelques corps protéiques (contenant la  $\gamma$ -zéine riche en lysine), la plus grande partie des corps protéiques isolés n'était pas immunomarkuée avec des anticorps  $\alpha$ P-L parce qu'ils correspondaient à des cellules d'endosperme non transformées. Pour déterminer si la  $\gamma$ -zéine riche en lysine était co-  
15 localisée avec les polypeptides  $\alpha$ -zéines et  $\gamma$ -zéines, un double marquage en immuno-électromicroscopie a été réalisé sur des corps protéiques isolés en utilisant les anticorps  $\alpha$ Z et  $\alpha$ P-L (Fig. 8B) et  $\alpha$ G2 et  $\alpha$ P-L (Fig. 8C et D). La figure 8B montre une micrographie de la section transversale de deux corps protéiques marqués avec l'anticorps  $\alpha$ P-L  
20 (particule d'or de 15 nm) et avec l'anticorps  $\alpha$ Z (particule d'or de 5 nm). Le résultat de l'immunocoloration a montré que la protéine P20 $\gamma$ Z est accumulée dans les corps protéiques et co-localisée avec l' $\alpha$ -zéine (voir l'étendue du marquage de l' $\alpha$ -zéine sur toute la surface du corps protéique). Par ailleurs, des sections tangentielles (Fig. 8B, voir les  
25 flèches) et transversales (Fig. 8D) des corps protéiques ont été incubées avec l'anticorps  $\alpha$ P-L (particule d'or 15 nm) et avec l'anticorps  $\alpha$ G2 (particule d'or de 5 nm). Dans les deux cas, la protéine P20 $\gamma$ Z était co-localisée avec les polypeptides  $\gamma$ -zéines. On a pu noter que les sections tangentielles du corps protéique (Fig. 8A, C, voir les flèches) étaient





facilement distinguées des sections transversales du corps protéique dans la mesure où les premiers présentaient une densité plus forte en électrons et que le marquage de la  $\gamma$ -zéine était étendu sur toute la section. Au contraire les sections transversales présentaient une densité inférieure et le marquage de la  $\gamma$ -zéine était localisé sur la membrane entourant le corps protéique. Dans les deux cas, la localisation du marquage de la  $\gamma$ -zéine riche en lysine suivait celle de la  $\gamma$ -zéine endogène.

**B) Préparation de plantes génétiquement modifiées, exprimant des  $\gamma$ -zéines riches en lysine**

**1) Obtention et utilisation de cal de maïs comme cible pour la transformation génétique.**

La transformation génétique du maïs, quelle que soit la méthode employée (electroporation; Agrobacterium, microfibres, canon à particules) requiert généralement l'utilisation de cellules indifférenciées en divisions rapides ayant conservé une aptitude à la régénération de plantes entières. Ce type de cellules compose le cal friable embryogène (dit de type II) de maïs.

Ces cals sont obtenus à partir d'embryons immatures de génotype Hi II ou (A188 x B73) selon la méthode et sur les milieux décrits par Armstrong (Maize Handbook ; (1994) M. Freeling, V. Walbot Eds ; pp 665-671). Les cals ainsi obtenus sont multipliés et maintenus par repiquages successifs tous les quinze jours sur le milieu d'initiation.

Des plantules sont ensuite régénérées à partir de ces cals en modifiant l'équilibre hormonal et osmotique des cellules selon la méthode décrite par Vain et al. (Plant Cell Tissue and Organ Culture (1989 18 : 143-



151). Ces plantes sont ensuite acclimatées en serre où elles peuvent être croisées ou autofécondées.

2) Utilisation du canon à particules pour la transformation  
5 génétique du maïs.

Le paragraphe précédent décrit l'obtention et la régénération des lignées cellulaires nécessaires à la transformation ; on décrit ici une méthode de transformation génétique conduisant à l'intégration stable des  
10 gènes modifiés dans le génome de la plante. Cette méthode repose sur l'utilisation d'un canon à particules identique à celui décrit par J. Finer (Plant Cell Report (1992) 11 :323-328) ; les cellules cibles sont des fragments des cals décrits dans le paragraphe 1. Ces fragments d'une surface de 10 à 20 mm<sup>2</sup> ont été disposés, 4h avant bombardement, à  
15 raison de 16 fragments par boîte au centre d'une boîte de petri contenant un milieu de culture identique au milieu d'initiation, additionné de 0,2 M de mannitol + 0,2M de sorbitol. Les plasmides portant les gènes à introduire sont purifiés sur colonne Qiagen® en suivant les instructions du fabricant. Ils sont ensuite précipités sur des particules de tungsten (M10) en suivant  
20 le protocole décrit par Klein (Nature (1987) 327 :70-73). Les particules ainsi enrobées sont projetées vers les cellules cibles à l'aide du canon et selon le protocole décrits par J. Finer (Plant Cell Report (1992) 11 : 323-328).

Les boîtes de cals ainsi bombardées sont ensuite scellées à l'aide  
25 de Scellofrais® puis cultivées à l'obscurité à 27°C. Le premier repiquage a lieu 24h après, puis tous les quinze jours pendant 3 mois sur milieu identique au milieu d'initiation additionné d'un agent sélectif dont la nature et la concentration peuvent varier selon le gène utilisé (voir paragraphe 3). Les agents sélectifs utilisables consistent généralement en composés



actifs de certains herbicides (Basta®, Round up®) ou certains antibiotiques (Hygromycine, Kanamycine...).

On obtient après 3 mois ou parfois plus tôt, des cals dont la croissance n'est pas inhibée par l'agent sélectif, habituellement et majoritairement composés de cellules résultant de la division d'une cellule ayant intégré dans son patrimoine génétique une ou plusieurs copies du gène de sélection. La fréquence d'obtention de tels cals est d'environ 0,8 cal par boîte bombardée.

Ces cals sont identifiés, individualisés, amplifiés puis cultivés de façon à régénérer des plantules (cf. paragraphe 1). Afin d'éviter toute interférence avec des cellules non transformées toutes ces opérations sont menées sur des milieux de culture contenant l'agent sélectif.

Les plantes ainsi régénérées sont acclimatées puis cultivées en serre où elles peuvent être croisées ou autofécondées.

15

### 3) Utilisation du gène Bar pour la production plantes de maïs génétiquement modifiées ayant incorporé et exprimant le gène H45yZ

Le gène Bar de *Streptomyces hygroscopicus* code pour une phosphotricine acyl transférase (PAT) qui inactive par acétylation la phosphotricine-molécule active de l'herbicide Basta®. Les cellules portant ce gène sont donc rendues résistantes à cet herbicide et peuvent être sélectionnées par son intermédiaire. Pour la transformation des céréales la séquence codante du gène Bar est sous le contrôle d'une région régulatrice permettant une expression forte et constitutive dans les cellules végétales. Une telle région est avantageusement constituée par le promoteur et le premier intron du gène Actine de riz tels que décrits par Mc Elroy (Mol. Gen. Genet. (1991) 231 : 150-160).

Ce gène chimérique est cloné sur un plasmide permettant son amplification par *Escherichia coli*. Ce plasmide (pDM 302 Cao (Plant Cell

30



Report (1992) 11:586-591) après amplification puis purification sur colonne Qiagen® peut être utilisé en transformation génétique du maïs en utilisant par exemple la méthode décrite dans l'exemple précédent. Dans ce cas les milieux de culture destinés à la sélection des cellules transformées sont  
5 additionnés de 2 mg/L de Phosphinotricine.

Pour l'introduction du gène H45yZ une technique dite de cotransformation est avantageusement utilisée : les gènes de sélection (Bar) et d'intérêt (H45yZ) sont portés par des plasmides indépendants. Dans le cas de l'utilisation du canon à particules on procède à une  
10 coprécipitation des plasmides sur les particules de tungstène, la quantité totale d'ADN précipité sur les particules restant identique à ce qu'elle est dans le protocole standard (5µg d'ADN pour 2,5 mg de particules) chaque plasmide représentant environ la moitié du poids total d'ADN utilisé.

L'expérience montre qu'avec cette méthode, la cointégration des  
15 plasmides dans les cellules végétales est l'évènement le plus fréquent (de l'ordre de 90%) c'est-à-dire que pratiquement chaque plante ayant intégré le gène Bar et sélectionnée par son intermédiaire portera aussi le gène H45yZ. Le niveau de coexpression (pourcentage de plantes sélectionnées exprimant le gène H45yZ) est habituellement de l'ordre de 70%.

20 Les gènes ainsi introduits sont généralement liés au sens génétique, le gène H45yZ peut ainsi avantageusement être suivi dans les descendance grâce à la résistance à l'herbicide qui lui est étroitement associée.

La quantité de protéine modifiée est déterminée par les méthodes  
25 décrites dans l'exemple A, notamment par immunoblotting sur extraits protéiques de grains de maïs immatures ou matures, prélevés en pool sur des plantes résistantes au BASTA.





4) Exemple explicitant l'étape d'introduction des transgènes et notamment du gène codant pour la H45 $\gamma$ Z, pour la modification du phénotype opaque 2 du maïs

5 Amélioration de maïs opaque 2 par l'introgression du gène de  $\gamma$ -zéine enrichie en lysine.

Les plantes transformées décrites dans l'exemple précédent, présentent à la fois, une résistance au Basta et l'expression d'une  $\gamma$ -zéine enrichie en lysine sont utilisées. Leur pollen est utilisé pour féconder les  
10 plantes de maïs opaque 2 de la lignée W64Ao2 qui ne contient qu'un seul gène  $\gamma$ -zéine. Cette lignée a été obtenue auprès du Maize Stock Center. Les plantes de ces descendance F1 sont sélectionnées pour leur résistance au Basta et ensuite autofécondées. Les graines F2 ainsi produites sont analysées pour le phénotype opaque sur une table  
15 lumineuse et les grains opaque ou les grains vitreux sont semés et évalués pour la résistance au Basta. Dans le cas où tous les grains opaques sont sensibles au Basta, la démonstration est faite que l'introduction du gène  $\gamma$ -zéine enrichi en lysine dans la plante considérée, complémente le phénotype opaque 2.

20 Dans un deuxième temps, dans ces plantes Basta résistantes, des individus de génotype o2/o2 et ne possédant qu'un seul gène  $\gamma$ -zéine sur le chromosome 7 sont sélectionnés à l'aide des sondes moléculaires codant pour le gène Opaque 2 et pour la  $\gamma$ -zéine. Ces dernières révélant des fragments de restriction polymorphes et seuls les individus présentant  
25 le patron type de la lignée W64o2 sont retenus (Lopes M.A. et al, 1995, Mol. Gen. Genet. 19 : 247, 603-613).

Ces individus ont un contenu en lysine qui est en moyenne équivalent à celui du maïs o2, voire supérieur. A partir de ces individus, toute introgression dans des variétés ELITE, du caractère « forte teneur en



lysine » est observée via la détermination de la résistance au BASTA et de la présence de l'allèle  $\alpha 2$  détectée par RFLP.

5) Expression de  $\gamma$ -zéines enrichies en lysine dans *Arabidopsis thaliana*.

5

Pour obtenir une transformation stable, les constructions plasmidiques P20 $\gamma$ Z et pH30 $\gamma$ Z clonées dans le plasmide Bluescript KS (-) ont été insérées sous forme de fragments HincII/XbaI dans le vecteur binaire pBin19 (Bevan, M. Nucl. Acids Res. 12: 8711-8721 (1984)),  
10 contenant le promoteur 35S du virus de la mosaïque du chou-fleur (CaMV) et les signaux de formation de l'extrémité 3' et de polyadénylation du gène de l'octopine synthétase (ocs). Les nouveaux plasmides ont été appelés p19P20 $\gamma$ Z et p19H30 $\gamma$ Z (figure 12).

Les vecteurs binaires contenant les séquences codant pour les  
15 protéines P20 $\gamma$ Z et H30 $\gamma$ Z (p19P20 $\gamma$ Z et p19H30 $\gamma$ Z) ont été transférés dans la souche LBA4404 d'*Agrobacterium tumefaciens*. Des plantes *Arabidopsis* ecotype RLD ont été transformées selon la méthode décrite par Valvekens, D., Van Montagu, M. et Van Lijsebettens, M. ((1988) Proc. Natl.Acad.Sci.USA 85: 5536-5540). Pour chaque construction, 10 plantes  
20 transgéniques ont été criblées par analyse en immunoblot en utilisant un antisérum obtenu contre la  $\gamma$ -zéine ( $\alpha$ G2, Ludevid et al. 1985). Les plantes contenant les taux les plus élevés (correspondant à environ 0,1% de la quantité totale de protéines présentes dans les feuilles d'*Arabidopsis*) de produits transgéniques dans la génération F1, ont été choisies pour obtenir  
25 la génération F2. Ces plantes ont également été sélectionnées pour l'expression de la protéine recherchée.

Des plantes transgéniques entières, sélectionnées dans un milieu contenant de la kanamycine ont été homogénéisées dans de l'azote liquide. Les protéines transgéniques ont été extraites sélectivement avec



une solution contenant de l'éthanol/ de l'acide chlorhydrique HCl 0,125 N dans une proportion de 3:1 (v/v), avec 5% de mercaptoéthanol et d'inhibiteurs de protéase. Les protéines extraites avec cette solution ont été précipitées dans 5 volumes d'acétone et analysées par SDS-PAGE et immunoblotting. Les extraits protéiques des plantes non transgéniques ont été utilisés en tant que contrôles. Les protéines résultant de l'insertion des séquences K(P-K)<sub>4</sub> dans la  $\gamma$ -zéine ont été exprimées correctement dans les plantes Arabidopsis thaliana en utilisant le promoteur constitutif 35 S du CaMV. Sur les immunoblots, les anticorps  $\alpha$ G2 et  $\alpha$ PL reconnaissaient des bandes d'électrophorèse correspondant aux protéines P20 $\gamma$ Z et H30 $\gamma$ Z. Ces bandes migraient avec des masses moléculaires apparentes conformes à celles que l'on avait préalablement observées dans les expériences de traduction / translocation in vitro (30 kD et 26 kD respectivement). Comme on l'a observé dans les plantes transgéniques Arabidopsis exprimant la  $\gamma$ -zéine (Geli et al. Plant Cell 6: 1911-1922 (1994)), les protéines P20 $\gamma$ Z et H30 $\gamma$ Z migraient sous la forme de deux bandes d'électrophorèse à savoir les bandes correspondant à 36 et 30 kD pour P20 $\gamma$ Z et les bandes correspondant à 32 et 26 kD pour H30 $\gamma$ Z. Les bandes supérieures pourraient correspondre à des produits ayant subi des modifications post-traductionnelles. Une telle modification post-traductionnelle n'a pas été détectée dans les endospermes de maïs transformés. Ce résultat suggère que la modification apparaîtrait lorsque ces protéines sont exprimées dans un système hétérologue tel que Arabidopsis thaliana.

25

6) Expression de  $\gamma$ -zéines enrichies en lysine recombinantes dans du maïs.

Méthode:



Après obtention des plantes transgéniques, on les croise avec une lignée mâle non transformée. Par conséquent, 50% des grains récoltés dans ce cas d'insertion unilocus seront transgéniques. Dans le but d'analyser les  $\gamma$ -zéines enrichies en lysine dans les plantes transgéniques, les protéines sont extraites à partir de 6 grains par transformant.

Les endospermes sont disséqués en retirant des grains les embryons et les pericarpes. Les endospermes sont broyés et 50mg de farine sont utilisés pour l'extraction sélective. Préalablement les  $\alpha$ -zéines sont extraites par 3 traitements avec de l'éthanol à 70%. Après centrifugation, l'éthanol est évaporé sous vide. Les protéines insolubles dans l'éthanol (principalement les  $\gamma$ -zéines et les  $\gamma$ -zéines enrichies en lysine) sont extraites du culot avec un tampon de Laemli contenant 10% de mercaptoéthanol (100 $\mu$ l de tampon pour 10mg de farine). Les protéines totales sont ensuite analysées par coloration à l'argent (Morrissey J.H. 1981, Ann. Biochem. vol. 117, p. 307-310). Les  $\gamma$ -zéines et les  $\gamma$ -zéines enrichies en lysine sont analysées par immunoblot en utilisant respectivement les anticorps  $\alpha$ G2 (dilution 1/2000) et  $\alpha$ PL (dilution 1/500). Un anticorps anti-anticorps de lapin conjugué à la phosphatase alcaline est utilisé comme anticorps secondaire dans l'immunoblot. Les extraits sont dilués pour permettre leur chargement sur SDS-page selon la méthode analytique utilisée.

#### Résultats:

#### Accumulation de $\gamma$ -zéines enrichies en lysine dans les plants 20 $\gamma$ Z et 45 $\gamma$ Z de maïs transgénique;

La figure 13 montre l'immunoblot des extraits de protéines révélées avec l'antisérum  $\alpha$ PL (A,B) et les protéines totales après coloration à l'argent (C). Comme on peut le voir dans la fig13A, 6 plants transgéniques 20 $\gamma$ Z ont été testés avec l'antisérum  $\alpha$ PL et de la  $\gamma$ -zéine enrichie en lysine est exprimée dans les plants transgéniques A1 et D2 (pistes 1 et 6





respectivement). Dans le plant C1 (piste 4), seules des traces de  $\gamma$ -zéine enrichie en lysine sont observées. Lorsque les extraits des 6 plants transgéniques 45 $\gamma$ Z sont chargés sur le gel et marqués avec l'antisérum  $\alpha$ PL (Fig 13B), on observe une forte réaction avec l'anticorps pour les transformants B1 et C1 (pistes 3 et 4 respectivement).

Il faut noter que la réaction avec l'anticorps dans les extraits des endospermes des plants 45 $\gamma$ Z B1 et C1 est plus forte que dans les extraits des plants 20 $\gamma$ Z A1 et D2. Ce résultat est confirmé après coloration à l'argent des gels (fig 13C) où les 2 types de  $\gamma$ -zéine : endogène et enrichie en lysine, sont colorés. La  $\gamma$ -zéine enrichie en lysine a un poids moléculaire apparent de 30kDa et la  $\gamma$ -zéine endogène de 28 kDa. L'expression de  $\gamma$ -zéine enrichie en lysine est plus faible dans le plant 45 $\gamma$ Z B1 que dans C1 (pistes 2 et 3 respectivement). Dans les pistes 1 à 6 de la Fig 13C, une dilution identique des protéines des extraits d'endospermes des plants 45 $\gamma$ Z et 20 $\gamma$ Z avait été déposée sur le gel. A cette dilution, l'expression de la  $\gamma$ -zéine enrichie en lysine est détectée seulement dans les endospermes des plants 45 $\gamma$ Z B1 et C1. Cependant, lorsqu'un extrait plus important (40  $\mu$ g de protéines par piste) de protéines 20 $\gamma$ Z est chargé sur le gel, une légère bande (voir flèche), correspondant à la  $\gamma$ -zéine enrichie en lysine est détectée dans les endospermes 20 $\gamma$ Z A1 et D2. Ce résultat indique que les plants 45 $\gamma$ Z accumulent beaucoup plus de protéines selon l'invention que les plants 20 $\gamma$ Z. Ceci est probablement dû aux activités différentes des promoteurs. Les plantes 45 $\gamma$ Z ont été transformées avec un construit contenant la  $\gamma$ -zéine selon l'invention sous le contrôle du promoteur  $\gamma$ -zéine (1,7 kb) alors que les plantes 20 $\gamma$ Z ont été transformées avec la même séquence codante mais sous le contrôle du promoteur 35S du CaMV. La coloration à l'argent est une technique générale de coloration des protéines, mais la forte couleur marron est surtout observée en présence de protéines basiques. Comme les  $\alpha$ -zéines



ont été extraites de la farine, comme décrit ci-dessus, elles sont absentes lors de l'analyse en SDS-page de la Fig 13C.

### Ségrégation

Puisqu'il y a une ségrégation de 50% dans toutes les plantes  
5 transgéniques dans le cas d'un locus unique, une analyse grain par grain a été réalisée afin de quantifier le taux de  $\gamma$ -zéine enrichie en lysine dans chaque transformant. L'analyse a été faite uniquement avec les transformants 45yZ qui présentaient le plus fort taux d'expression. La figure 14 montre la coloration à l'argent (A) et l'immunoblot avec  $\alpha$ PL (B)  
10 de 5 endospermes différents de 45yZ B1 et C1. La faible bande électrophorétique présente sur toutes les pistes (voir A, pistes 1 à 10) correspond à la  $\gamma$ -zéine endogène. Comme on peut le voir dans la Fig 14 A, 2 des 5 endospermes accumulent de la  $\gamma$ -zéine enrichie en lysine (voir pistes 3,4 et 6,7). Par conséquent, environ la moitié des graines  
15 accumulent des quantités significatives de protéines enrichies en lysine. Si on prend en compte le fait que des quantités identiques de protéines ont été déposées sur le gel, on observe que le transformant 45yZ C1 a accumulé plus de  $\gamma$ -zéine enrichie en lysine que B1. En fait ce résultat est en accord avec ce qui est observé lors de la coloration à l'argent des  
20 extraits mélangés d'endosperme (Fig 13 C, piste 3). La preuve de la présence de  $\gamma$ -zéine enrichie en lysine dans ces endospermes est soulignée dans la Fig 14B. L'immunoblot utilisant l'antisérum  $\alpha$ PL montre que 2 extraits d'endosperme de 45yZ B1 (pistes 3 et 4) et de 45yZ C1 (pistes 4 et 6) accumulent de la  $\gamma$ -zéine enrichie en lysine. Pour confirmer  
25 le pourcentage de grains transgéniques, 10 nouveaux grains de transformant 45yZ C1 ont été analysés par immunoblot et utilisation de l'antisérum  $\alpha$ PL (Fig 15 A). Comme attendu, environ la moitié des graines transgéniques sont détectées. Une bande immunoréactive a été observée dans les extraits d'endosperme (pistes 1,2,9 et 10).



Estimation de la quantité de  $\gamma$ -zéïnes enrichies en lysine dans les endospermes des transformants 45 $\gamma$ Z:

5  $\alpha$ G2 est un antisérum polyclonal qui reconnaît les  $\gamma$ -zéïnes endogènes et les  $\gamma$ -zéïnes enrichies en lysine. La réactivité de cet antisérum avec les extraits d'endospermes 45 $\gamma$ Z C1 a été utilisée pour quantifier la quantité de  $\gamma$ -zéïnes enrichies en lysine selon l'invention dans les endospermes des plantes transformées.

10 La figure 15B présente l'immunoblot de 5 extraits protéiques correspondant à 5 endospermes 45 $\gamma$ Z C1 (pistes 1 à 5). Comme attendu, seuls 2 endospermes ont présenté un profil d'immunoréaction caractéristique des grains transgéniques. La bande supérieure de 30 kDa correspond à la  $\gamma$ -zéïne enrichie en lysine (flèches sur piste 1 et 2) et la bande inférieure à la  $\gamma$ -zéïne endogène. On peut noter que dans les  
15 endospermes des plantes non transgéniques la bande supérieure est absente (pistes 3, 4 et 5). De façon surprenante, il semble que le contenu en  $\gamma$ -zéïne endogène est plus faible dans les extraits des plantes transgéniques que dans les plantes non-transgéniques (voir flèches sur pistes 1 et 5).

20 A première vue, il a été observé que :

- i) dans les endospermes transgéniques 45 $\gamma$ Z C1, le ratio  $\gamma$ -zéïne enrichie en lysine/ $\gamma$ -zéïne endogène est de 7/3. Donc la quantité de protéine modifiée selon l'invention est au moins 2 fois plus importante que celle de la protéine endogène,
- 25 ii) la quantité de  $\gamma$ -zéïne endogène dans les endospermes non transgéniques (voir pistes 3, 4 et 5) était équivalente à celle de la  $\gamma$ -zéïne enrichie en lysine dans les endospermes transgéniques (voir pistes 1 et 2).



7) **Expression de  $\gamma$ -zéines enrichies en lysine recombinantes dans du blé.**

Comme dans l'exemple 6), il est possible de montrer la présence de  $\gamma$ -zéines enrichies en lysine selon l'invention dans du blé.

- 5 La transformation du blé peut notamment être effectuée selon la méthode décrite dans Weeks et al., 1993, Plant Physiol., vol 102 : pages 1077-1084 ou selon la méthode décrite dans EP709462.





## REVENDEICATIONS

1. Oligonucléotide comprenant au moins un enchaînement  
5 codant pour un polypeptide répondant à la formule  $(P-K)_n$ , dans laquelle :
  - n est un entier supérieur ou égal à 2,
  - P représente un résidu d'acide aminé Proline,
  - K représente un résidu d'acide aminé Lysine,
  - le signe "-" symbolise une liaison entre les deux résidus d'acides10 aminés et notamment une liaison de type peptidique, les n unités (P-K) étant liées entre elles également par de telles liaisons par exemple de type peptidique.
2. Oligonucléotide selon la revendication 1, comprenant un enchaînement codant pour un polypeptide répondant à la formule  $(P-K)_n$ ,  
15 dans laquelle n est un entier supérieur ou égal à 3, de préférence n est égal à 4, 5, 6, 7, 8, 9, 10 ou 15.
3. Oligonucléotide selon la revendication 1 ou la revendication 2, comprenant un enchaînement codant pour un polypeptide répondant à la formule  $(P-K)_n$ , dans laquelle la séquence des n unités (P-  
20 K) est interrompue par un ou plusieurs résidus d'acides aminés différents des résidus P ou K.
4. Oligonucléotide selon l'une quelconque des revendications 1 à 3, caractérisé en ce que l'enchaînement codant pour le polypeptide comprenant les n unités (P-K) est complété à son extrémité 5' et/ou à son  
25 extrémité 3', par un ou plusieurs codons, codant par exemple pour au moins un résidu lysine à l'extrémité N-terminale du polypeptide formé.
5. Oligonucléotide selon la revendication 4, caractérisé en ce qu'il code pour un polypeptide répondant à la formule  $(P-K)_n$ , à la formule  $K(P-K)_4$  ou à la formule  $2K(P-K)_4$ .



6. Séquence de nucléotides recombinant comprenant un enchaînement de nucléotides codant pour une protéine de réserve d'une plante, caractérisée en ce qu'elle comprend en outre un oligonucléotide selon l'une quelconque des revendications 1 à 5, inséré en un site de  
5 l'enchaînement de nucléotides choisi de façon telle que :

- l'expression de la séquence de nucléotides dans une cellule végétale déterminée permet d'obtenir une protéine de réserve modifiée localisée de façon identique ou similaire à la protéine de réserve normale qui serait exprimée dans les mêmes conditions dans la même cellule par  
10 l'enchaînement de nucléotides codant correspondant, et/ou

- la protéine de réserve modifiée codée par la séquence de nucléotides recombinant est reconnue immunologiquement par des anticorps produits contre la protéine de réserve normale correspondante.

7. Séquence de nucléotides selon la revendication 6,  
15 caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve naturellement pauvre en lysine.

8. Séquence de nucléotides selon la revendication 7, caractérisée en ce que l'enchaînement de nucléotides codant qu'elle  
20 comprend, code pour une protéine de réserve naturellement produite par une plante utilisable en alimentation animale ou humaine.

9. Séquence de nucléotides selon la revendication 7 ou la revendication 8, caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve naturellement  
25 produite par une plante de la famille des céréales.

10. Séquence de nucléotides selon la revendication 7 ou la revendication 8, caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve naturellement produite par une plante de la famille des légumineuses ou des crucifères.



11. Séquence de nucléotides selon la revendication 9, caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve de maïs.

12. Séquence de nucléotides selon la revendication 11,  
5 caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve de la famille des zéines.

13. Séquence de nucléotides selon la revendication 12, caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve est la  $\gamma$ -zéine du maïs.

10 14. Séquence de nucléotides selon la revendication 13, caractérisée en ce que l'enchaînement de nucléotides codant pour la  $\gamma$ -zéine de maïs qu'elle contient, répond à la séquence présentée sur la Figure 9.

15 15. Séquence de nucléotides selon la revendication 7 ou la revendication 8, caractérisée en ce que l'enchaînement de nucléotides codant qu'elle comprend, code pour une protéine de réserve d'une plante choisie parmi les suivantes : le soja, le tournesol, le tabac, le blé, l'avoine, la luzerne, le riz, le colza, Arabidopsis thaliana.

20 16. Séquence de nucléotides selon la revendication 13 ou la revendication 14, caractérisée en ce que l'oligonucléotide est inséré à la place ou à la suite du domaine Pro-X ou dans le domaine Pro-X naturellement présent dans la  $\gamma$ -zéine du maïs.

25 17. Séquence de nucléotides recombinante, caractérisée en ce qu'elle comprend une séquence de nucléotides selon l'une quelconque des revendications 1 à 16 sous le contrôle d'un promoteur d'expression.

18. Séquence de nucléotides recombinante selon la revendication 17, caractérisée en ce que le promoteur est un promoteur spécifique d'un tissu cellulaire déterminé, par exemple un promoteur



spécifique pour l'expression dans les graines, et/ou dans les feuilles de plantes.

19. Séquence de nucléotides selon la revendication 17, caractérisée en ce que le promoteur d'expression est celui de la  $\gamma$ -zéine du  
5 maïs.

20. Séquence de nucléotides selon la revendication 17, caractérisée en ce que le promoteur d'expression est le promoteur CaMV35S.

21. Séquence de nucléotides selon l'une quelconque des  
10 revendications 16 à 20, caractérisée en ce qu'elle code pour l'un des polypeptides P20 $\gamma$ Z ou PH45 $\gamma$ Z répondant respectivement aux séquences représentées aux figures 11 et 10.

22. Vecteur de clonage et/ou d'expression, caractérisé en ce qu'il comprend, en un site non essentiel pour sa réplication, une séquence de  
15 nucléotides selon l'une quelconque des revendications 1 à 20.

23. Vecteur de clonage et/ou d'expression, caractérisé en ce qu'il s'agit de l'un des plasmides pP20 $\gamma$ Z (CNCM N° I-1640), pH30 $\gamma$ Z ou pH45 $\gamma$ Z (CNCM N° I-1639).

24. Polypeptide codé par une séquence selon l'une quelconque  
20 des revendications 1 à 17.

25.  $\gamma$ -zéine de maïs modifiée riche en lysine, caractérisée en ce qu'elle est codée par la séquence de nucléotides selon la revendication 13 ou 14.

26.  $\gamma$ -zéine de maïs modifiée enrichie en lysine, caractérisée en  
25 ce que sa séquence d'acides aminés est modifiée par au moins un polypeptide répondant à la formule (P-K)<sub>n</sub> ou à la formule K-(P-K)<sub>n</sub>, dans laquelle :

- n est un entier supérieur ou égal à 2,
- P représente un résidu d'acide aminé Proline,





- K représente un résidu d'acide aminé Lysine,

- le signe "-" symbolise une liaison entre les deux résidus d'acides aminés et notamment une liaison de type peptidique, les n unités (P-K) étant liées par des liaisons notamment de type peptidique, ledit polypeptide répondant à la formule (P-K)<sub>n</sub> ou K-(P-K)<sub>n</sub> étant substitué à une séquence naturellement présente dans la  $\gamma$ -zéine de maïs normale ou étant inséré avec délétion d'un ou plusieurs acides aminés de la séquence d'acides aminés de la  $\gamma$ -zéine de maïs normale, ou étant ajouté dans la séquence d'acides aminés de la  $\gamma$ -zéine normale, le site d'insertion du polypeptide étant choisi de façon telle que :

- lorsque la  $\gamma$ -zéine modifiée riche en lysine est produite dans une cellule hôte, notamment une cellule végétale, elle est localisée de façon identique ou similaire à la  $\gamma$ -zéine de maïs normale qui serait produite dans les mêmes conditions, dans la même cellule hôte, et/ou,

- la  $\gamma$ -zéine de maïs modifiée est reconnue par des anticorps dirigés contre la  $\gamma$ -zéine normale de maïs.

27.  $\gamma$ -zéine de maïs modifiée selon la revendication 26, caractérisée en ce qu'il s'agit de la protéine P20 $\gamma$ Z ou de la protéine H30 $\gamma$ Z ou de la protéine H45 $\gamma$ Z.

28. Cellule hôte recombinante, caractérisée en ce qu'elle comprend une séquence de nucléotides selon l'une quelconque des revendications 1 à 17.

29. Cellule hôte selon la revendication 28, caractérisée en ce qu'il s'agit d'une bactérie, par exemple de E. coli ou Agrobacterium tumefaciens.

30. Cellule hôte selon la revendication 28, caractérisée en ce qu'il s'agit d'une cellule végétale.

31. Cellule hôte selon la revendication 30, caractérisée en ce qu'il s'agit d'une cellule de semences de plante.



32. Cellule hôte selon la revendication 31, caractérisée en ce qu'il s'agit d'une cellule d'endosperme de semences de maïs.

33. Cellule hôte selon la revendication 32, caractérisée en ce qu'elle contient, intégrée de façon stable dans son génome, une séquence  
5 de nucléotides selon la revendication 13.

34. Cellule hôte selon la revendication 32, caractérisée en ce qu'elle produit une  $\gamma$ -zéine de maïs modifiée riche en lysine selon la revendication 26 ou la revendication 27.

35. Cellule hôte selon la revendication 30, caractérisée en ce  
10 qu'il s'agit d'une cellule de soja, de tournesol, de tabac, de blé, d'avoine, de luzerne, de riz, de colza, d'Arabidopsis.

36. Semences produisant un polypeptide selon l'une quelconque des revendications 24 à 27.

37. Plante produisant un polypeptide selon l'une quelconque des  
15 revendications 24 à 27.

38. Plante selon la revendication 37, caractérisée en ce qu'il s'agit du maïs.

39. Semences obtenues à partir des plantes selon la revendication 37 ou 38.

20 40. Procédé de production de plantes ou de semences exprimant une protéine de réserve modifiée, caractérisé en ce qu'il comprend les étapes de :

a) transformation d'une cellule végétale, avec une séquence de nucléotides selon l'une quelconque des revendications 6 à 17, ou un  
25 vecteur selon l'une quelconque des revendications 22 ou 23, dans des conditions permettant l'expression de façon stable et fonctionnelle de la protéine de réserve modifiée codée par la séquence de nucléotides;



b) régénération de plantes à partir de la cellule de plante transformée de l'étape a), pour obtenir des plantes exprimant la protéine de réserve modifiée,

c) le cas échéant, obtention de semences à partir des plantes  
5 modifiées obtenues à l'étape b).

41. Procédé selon la revendication 40, caractérisé en ce que la plante est le maïs et la protéine de réserve est la  $\gamma$ -zéine.



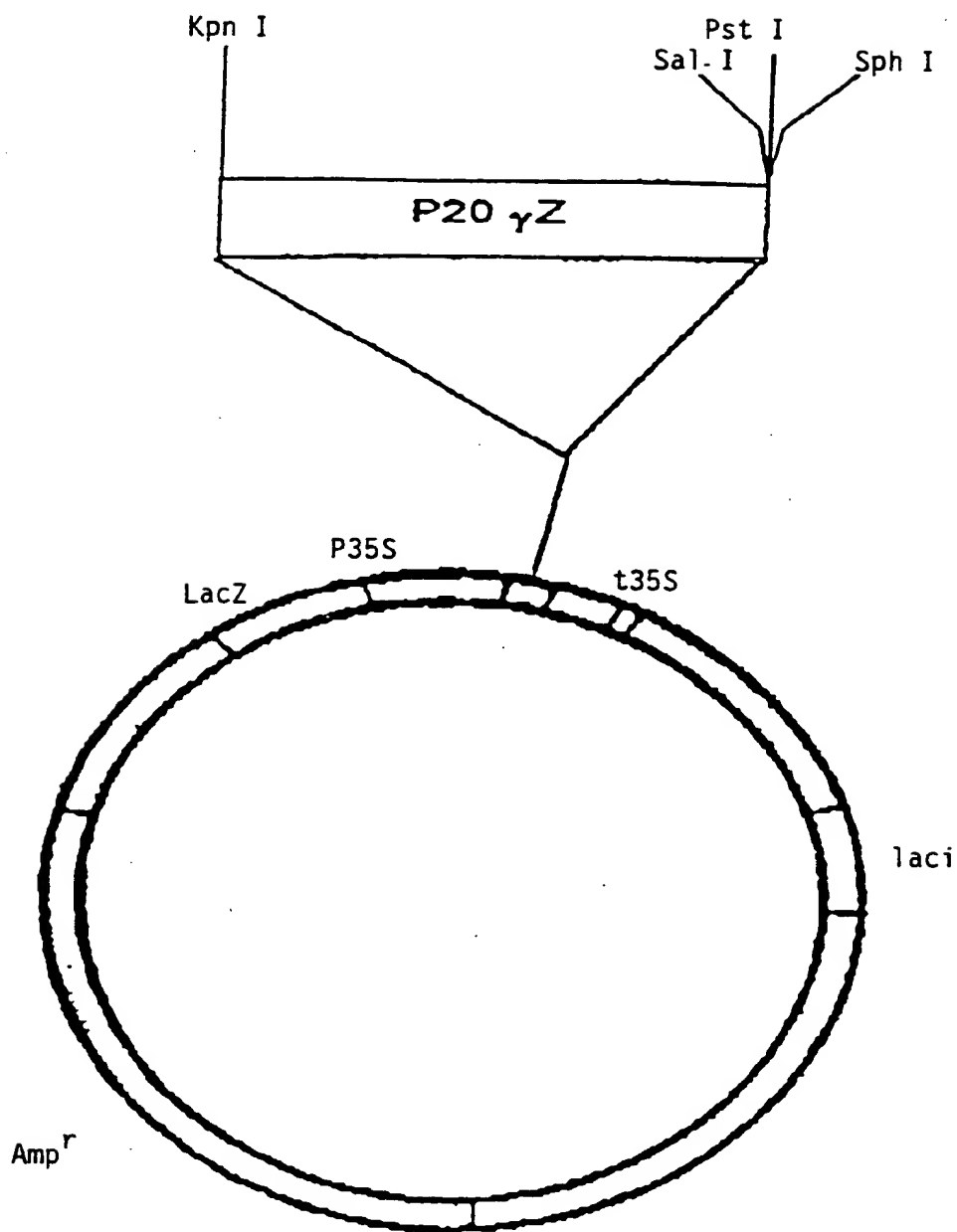
pP20 $\gamma$ Z

FIGURE 1





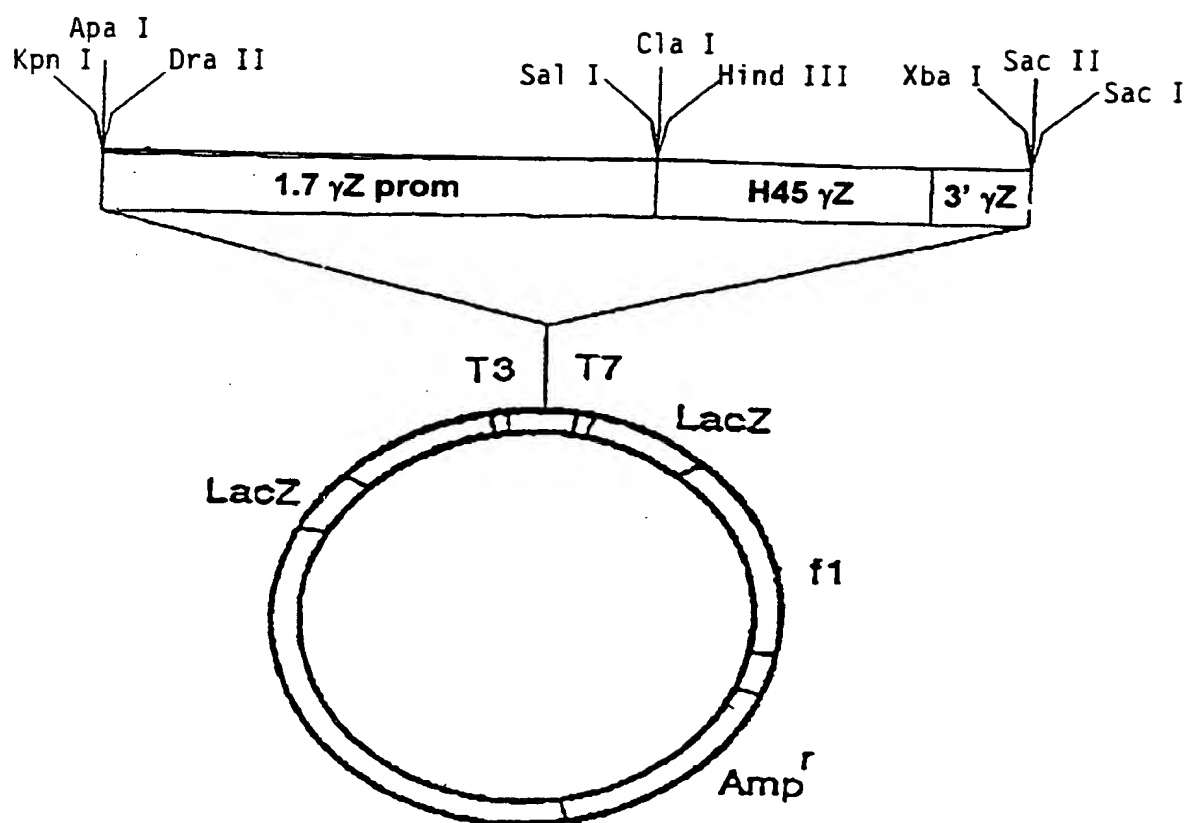
pH45 $\gamma$ Z

FIGURE 2



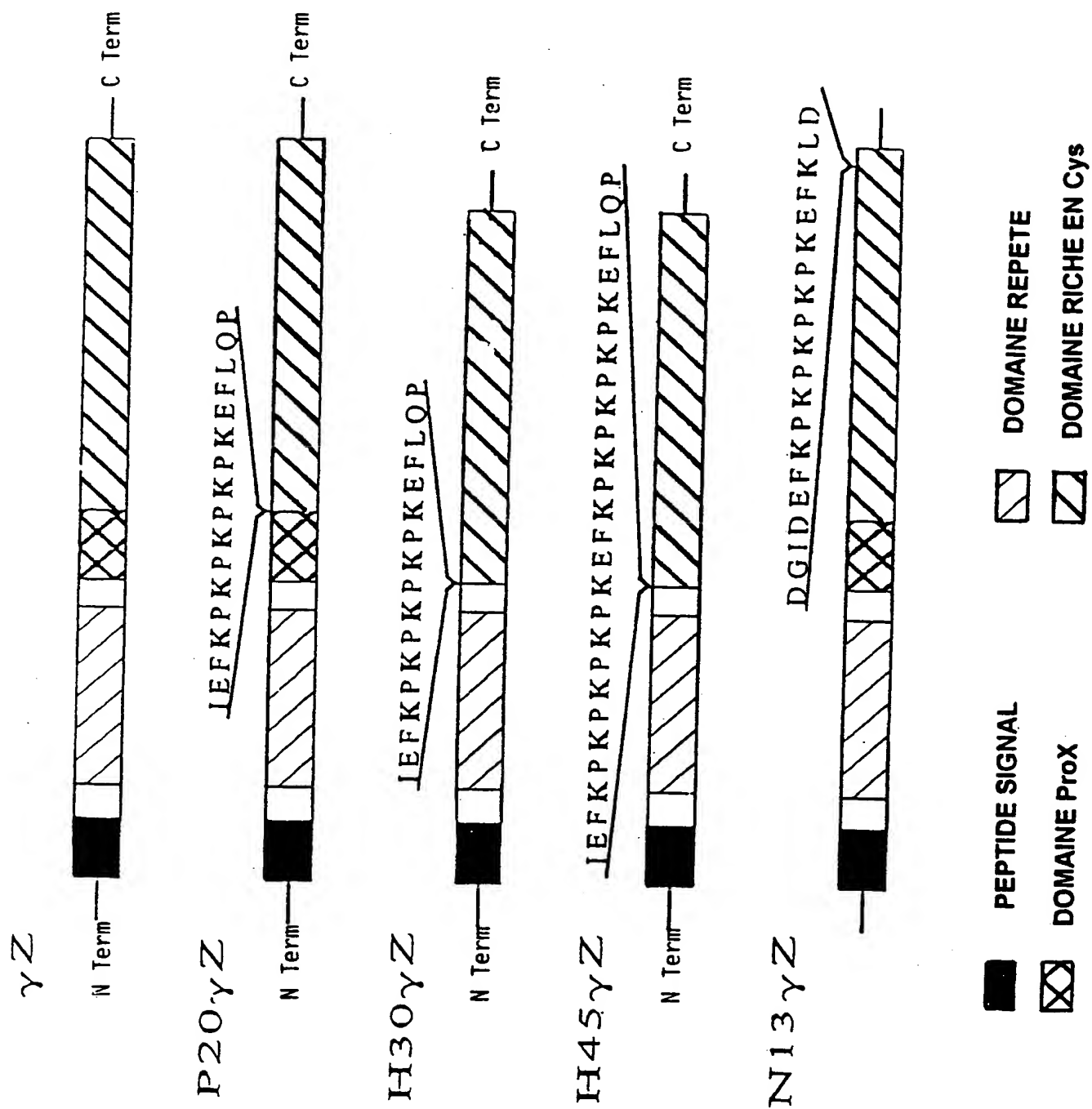


FIGURE 3



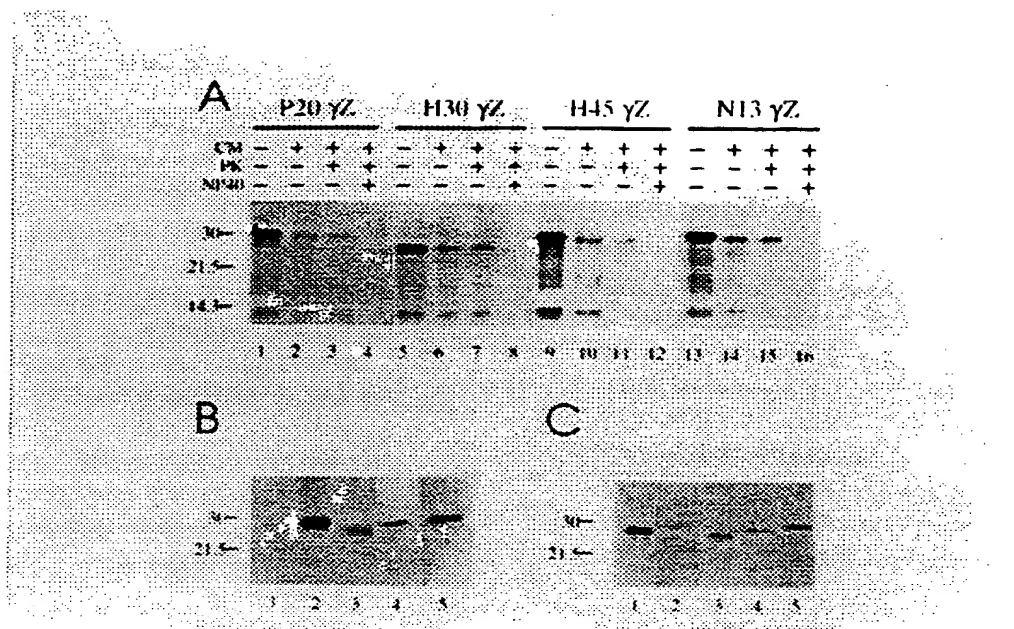


FIGURE 4

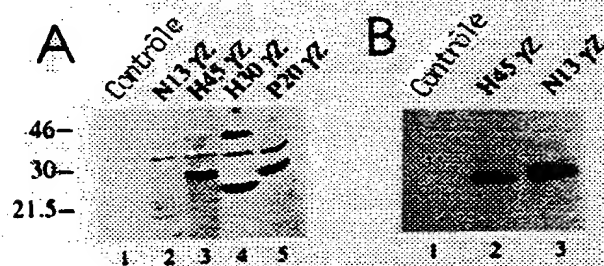


FIGURE 6



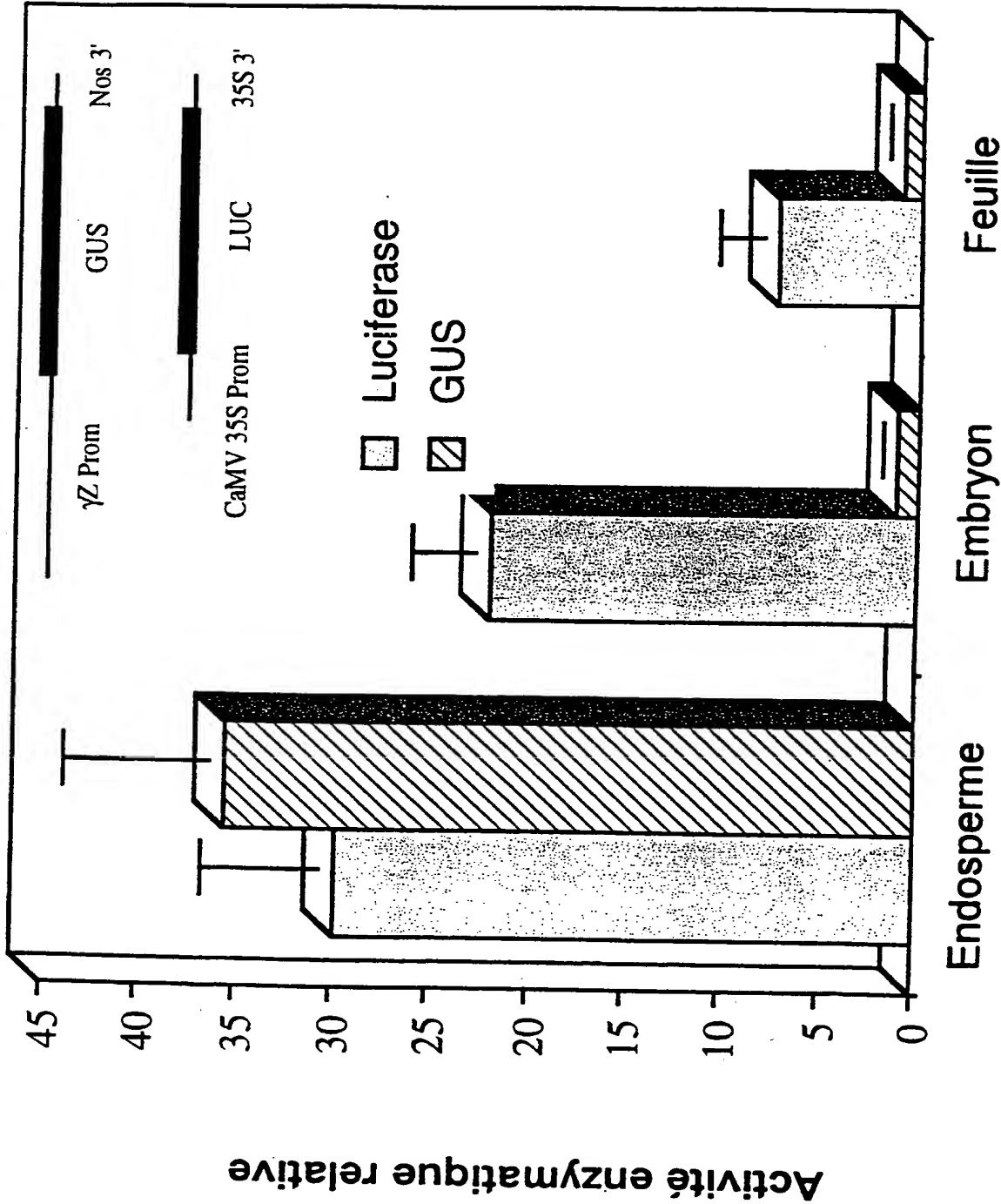


FIGURE 5





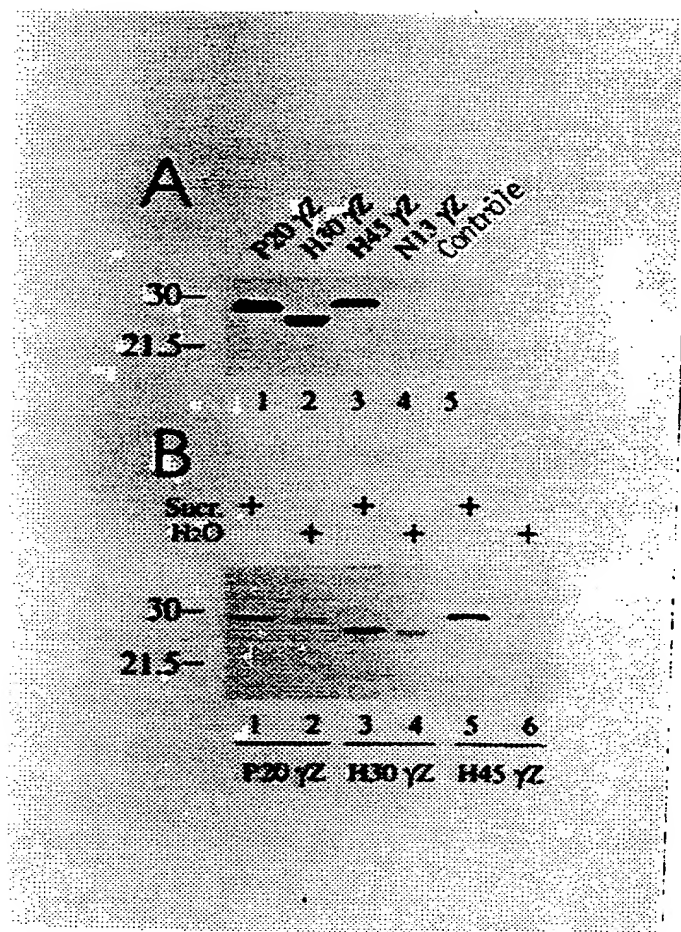


FIGURE 7



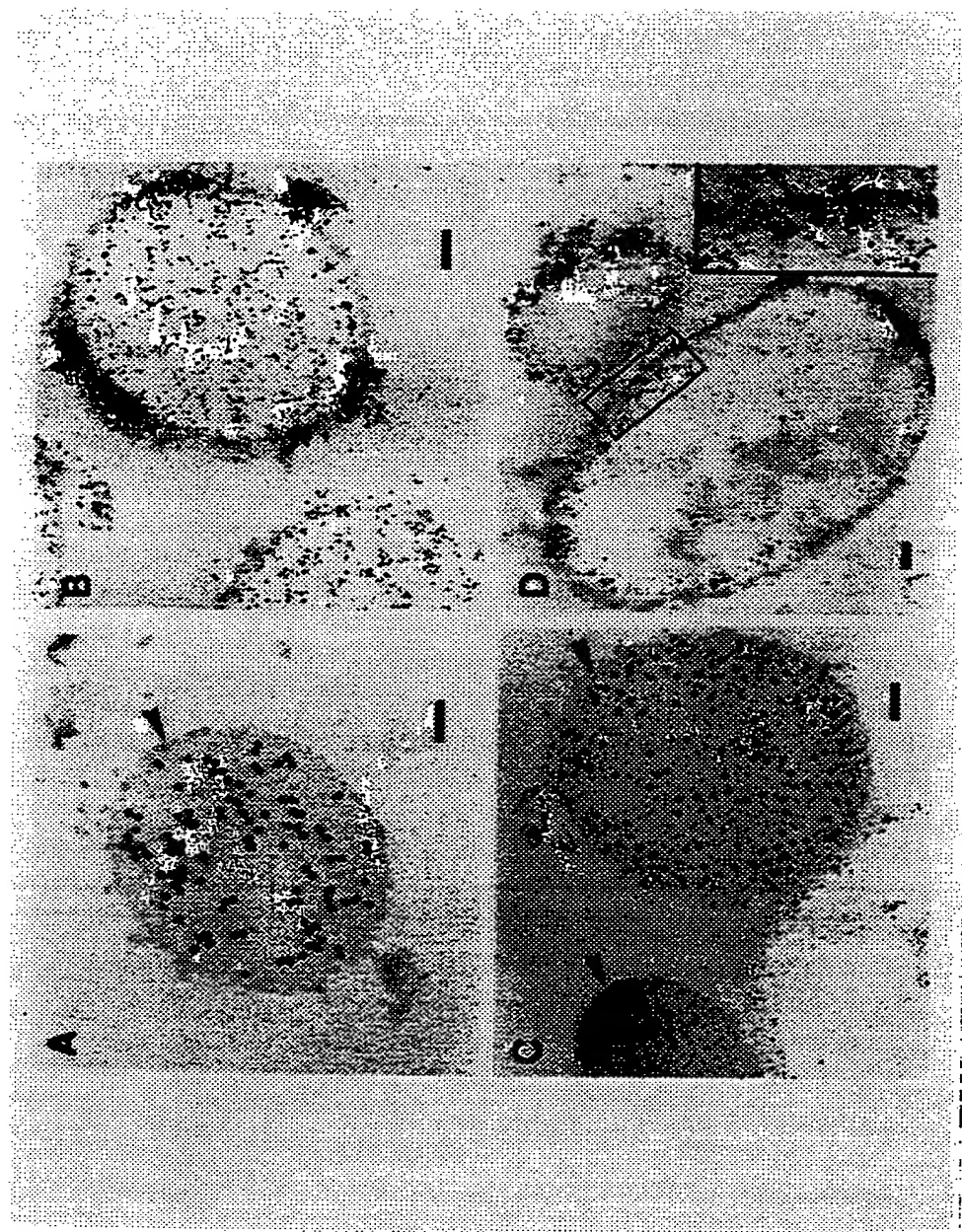


FIGURE 8



$\gamma$ -Zéine

10	20	30	40	50	60
ATGAGGGTGTGCTCGTTGCCCTCGCTCTCCTGGCTCTCGCTGCGAGCGCCACCTCCACG					
M R V L L V A L L A L A A S A T S T					
70	80	90	100	110	120
CATACAAGCGGCGGCTGCGGCTGCCAGCCACCGCCGCGGTTTCATCTACCGCCGCGGTTG					
H T S G G C G C Q P P P P V H L P P P V					
130	140	150	160	170	180
CATCTGCCACCTCCGGTTCACCTGCCACCTCCGGTGCATCTCCCACCGCCGGTCCACCTG					
H L P P P V H L P P P V H L P P P V H L					
190	200	210	220	230	240
CCGCCGCGGTTCCACCTGCCACCGCCGGTCCATGTGCCGCGCGGTTTCATCTGCCGCGG					
P P P V H L P P P V H V P P P V H L P P					
250	260	270	280	290	300
CCACCATGCCACTACCCTACTCAACCGCCCCGGCCTCAGCCTCATCCCCAGCCACACCCA					
P P C H Y P T Q P P R P Q P H P Q P H P					
310	320	330	340	350	360
TGCCCGTGCCAACAGCCGCATCCAAGCCCGTGCCAGCTGCAGGGAACCTGCGGCGTTGGC					
C P C Q Q P H P S P C Q L Q G T C G V G					
370	380	390	400	410	420
AGCACCCGATCCTGGGCCAGTGCCTCGAGTTTCTGAGGCATCAGTGCAGCCCGACGGCG					
S T P I L G Q C V E F L R H Q C S P T A					
430	440	450	460	470	480
ACGCCCTACTGCTCGCCTCAGTGCCAGTCGTTGCGGCAGCAGTGTGTCAGCAGCTCAGG					
T P Y C S P Q C Q S L R Q Q C C Q Q L R					
490	500	510	520	530	540
CAGGTGGAGCCGCAGCACCGGTACCAGGCGATCTTCGGCTTGGTCCTCAGTCCATCCTG					
Q V E P Q H R Y Q A I F G L V L Q S I L					
550	560	570	580	590	600
CAGCAGCAGCCGCAAAGCGGCCAGGTCGCGGGGCTGTTGGCGGCGCAGATAGCGCAGCAA					
Q Q Q P Q S G Q V A G L L A A Q I A Q Q					

FIGURE 9  
FEUILLE DE REMPLACEMENT (REGLE 26)



610 620 630 640 650 660  
| | | | |  
CTGACGGCGATGTGCGGCCTGCAGCAGCCGACTCCATGCCCCTACGCTGCTGCCGGCGGT  
L T A M C G L Q Q P T P C P Y A A A G G  
  
670  
|  
GTCCCCCACTGA  
V P H -

FIGURE 9 (SUITE)





H45  $\gamma$  Z

10	20	30	40	50	60
ATGAGGGTGTGCTCGTTGCCCTCGCTCTCCTGGCTCTCGCTGCGAGCGCCACCTCCACG					
M R V L L V A L A L L A A S A T S T					
70	80	90	100	110	120
CATACAAGCGGGCTGCGGCTGCCAGCCACCGCCGCCGGTTCATCTACCGCCGCCGGTG					
H T S G G C G C Q P P P P V H L P P P V					
130	140	150	160	170	180
CATCTGCCACCTCCGGTTCACCTGCCACCTCCGGTGATCTCCACCGCCGGTCCACCTG					
H L P P P V H L P P P V H L P P P V H L					
190	200	210	220	230	240
CCGCCGCCGGTCCACCTGCCACCGCCGGTTCATGTGCCGCCGCCGGTTCATCTGCCGCCG					
P P P V H L P P P V H V P P P V H L P P					
250	260	270	280	290	300
CCACCATGCCACTACCCTACTCAACCGCCCCGGATCGAATTCAAACCAAAGCCAAAGCCG					
P P C H Y P T Q P P R I E F K P K P K P					
310	320	330	340	350	360
AAGCCAAAAGAATTCAAACCAAAGCCAAAGCCGAAGCCAAAAGAATTCTGCAGCCCCTG					
K P K E F K P K P K P K P K E F L Q P L					
370	380	390	400	410	420
CAGGGAACCTGCGGCGTTGGCAGCACCCCGATCCTGGGCCAGTGCGTTCGAGTTTCTGAGG					
Q G T C G V G S T P I L G Q C V E F L R					
430	440	450	460	470	480
CATCAGTGACGCCCCGACGGCGACGCCCTACTGCTCGCCTCAGTGCCAGTCGTTGCGGCAG					
H Q C S P T A T P Y C S P Q C Q S L R Q					
490	500	510	520	530	540
CAGTGTTGCCAGCAGCTCAGGCAGGTGGAGCCGCAGCACCGGTACCAGGCGATCTTCGGC					
Q C C Q Q L R Q V E P Q H R Y Q A I F G					
550	560	570	580	590	600
TTGGTCCTCCAGTCCATCCTGCAGCAGCAGCCGCAAAGCGGCCAGGTTCGCGGGGCTGTG					
L V L Q S I L Q Q Q P Q S G Q V A G L L					
610	620	630	640	650	660
GCGGCGCAGATAGCGCAGCAACTGACGGCGATGTGCGGCCTGCAGCAGCCGACTCCATGC					
A A Q I A Q Q L T A M C G L Q Q P T P C					

FIGURE 10

FEUILLE DE REMPLACEMENT (REGLE 26)







P20  $\gamma$ Z

10 20 30 40 50 60  
| | | | | |  
ATGAGGGTGTGCTCGTTGCCCTCGCTCTCCTGGCTCTCGCTGCGAGCGCCACCTCCACG  
M R V L L V A L A L L A L A A S A T S T

70 80 90 100 110 120  
| | | | | |  
CATAACAAGCGGCGGCTGCGGCTGCCAGCCACCGCCGCGGTTTCATCTACCGCCGCGGTTG  
H T S G G C G C Q P P P P V H L P P P V

130 140 150 160 170 180  
| | | | | |  
CATCTGCCACCTCCGGTTCACCTGCCACCTCCGGTGCATCTCCACCGCCGGTCCACCTG  
H L P P P V H L P P P V H L P P P V H L

190 200 210 220 230 240  
| | | | | |  
CCGCCGCGGTTCCACCTGCCACCGCCGGTCCATGTGCCGCGCGGTTTCATCTGCCGCGG  
P P P V H L P P P V H V P P P V H L P P

250 260 270 280 290 300  
| | | | | |  
CCACCATGCCACTACCCTACTCAACCGCCCCGGCCTCAGCCTCATCCCCAGCCACACCCA  
P P C H Y P T Q P P R P Q P H P Q P H P

310 320 330 340 350 360  
| | | | | |  
TGCCCGTGCCAAACAGCCGCATCCAAGCCCGTGCCAGATCGAATTCAAACCAAAGCCAAAG  
C P C Q Q P H P S P C Q I E F K P K P K

370 380 390 400 410 420  
| | | | | |  
CCGAAGCCAAAAGAATTCCTGCAGCCCCTGCAGGGAACCTGCGGCGTTGGCAGCACCCCG  
P K P K E F L Q P L Q G T C G V G S T P

430 440 450 460 470 480  
| | | | | |  
ATCCTGGGCCAGTGCGTCGAGTTTCTGAGGCATCAGTGCAGCCCGACGGCGACGCCCTAC  
I L G Q C V E F L R H Q C S P T A T P Y

490 500 510 520 530 540  
| | | | | |  
TGCTCGCCTCAGTGCCAGTCGTTGCGGCAGCAGTGTGCGCAGCAGCTCAGGCAGGTGGAG  
C S P Q C Q S L R Q Q C C Q Q L R Q V E

550 560 570 580 590 600  
| | | | | |  
CCGCAGCACCGGTACCAGGCGATCTTCGGCTTGGTCCTCCAGTCCATCCTGCAGCAGCAG  
P Q H R Y Q A I F G L V L Q S I L Q Q Q

610 620 630 640 650 660  
| | | | | |  
CCGCAAAGCGGCCAGGTGCGCGGGGCTGTTGGCGGCGCAGATAGCGCAGCAACTGACGGCG  
P Q S G Q V A G L L A A Q I A Q Q L T A

FIGURE 11

FEUILLE DE REMPLACEMENT (REGLE 26)



670	680	690	700	710	720														
ATGTGCGGCCTGCAGCAGCCGACTCCATGCCCCCTACGCTGCTGCCGGCGGTGTCCCCCAC																			
M	C	G	L	Q	Q	P	T	P	C	P	Y	A	A	A	G	G	V	P	H

TGA

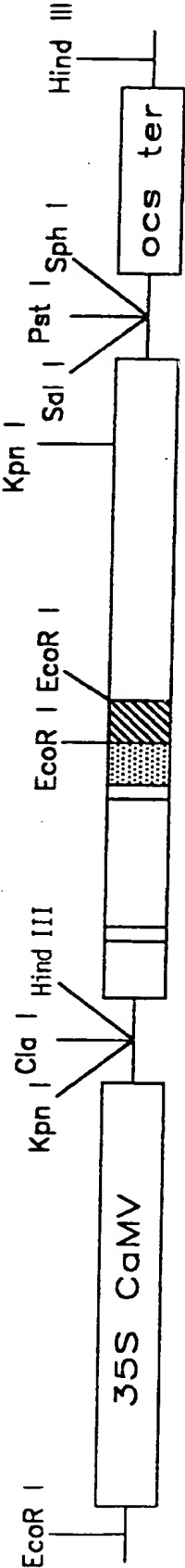
-

FIGURE 11 (SUITE)





pBin 19 P20γZ



pBin 19 H30γZ

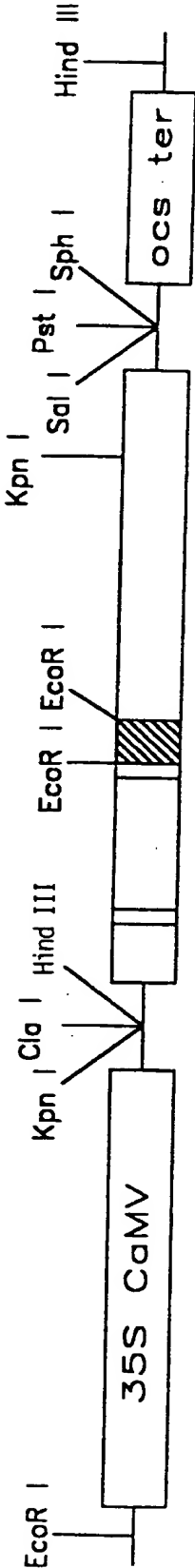


FIGURE 12



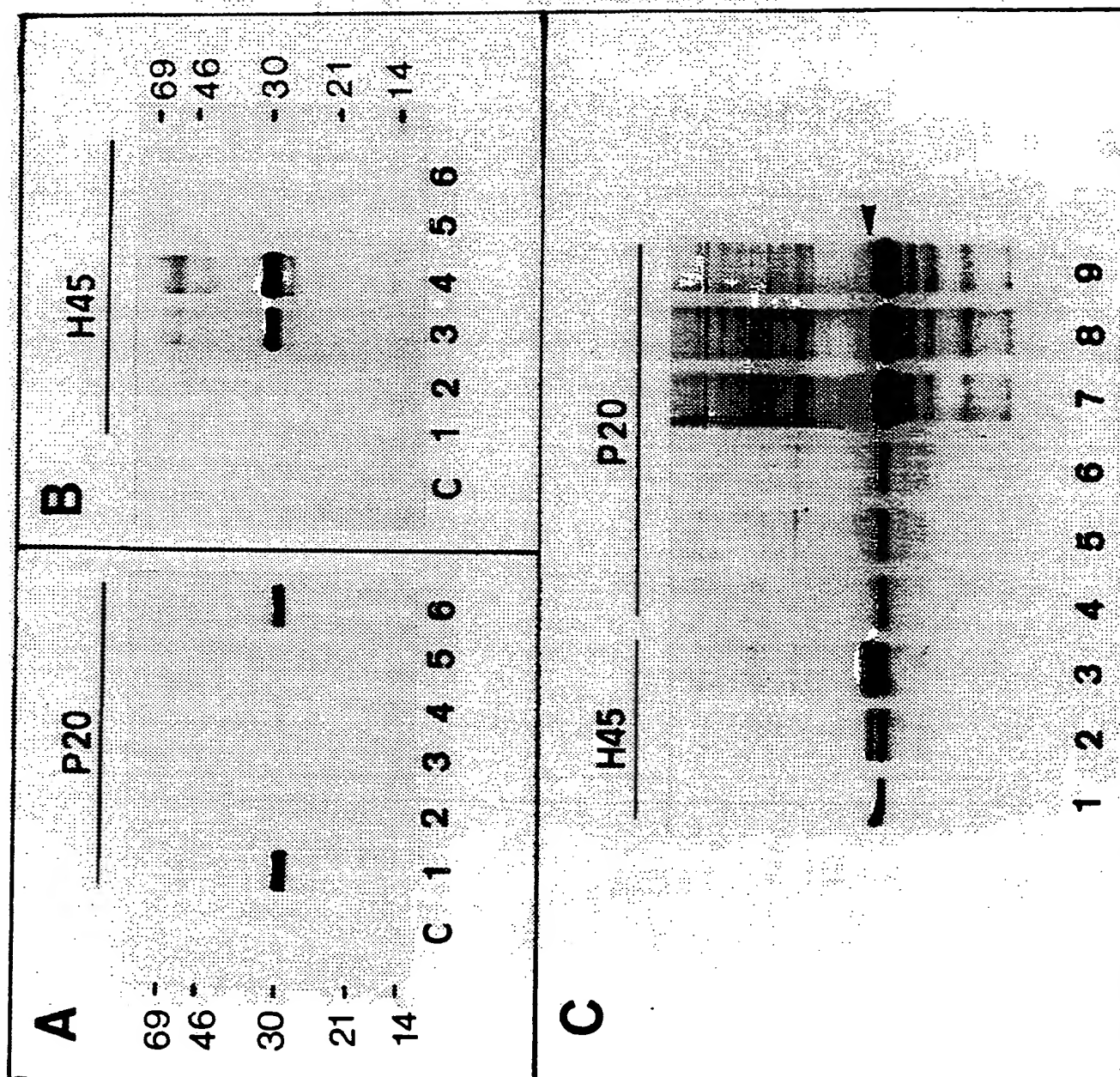


FIGURE 13



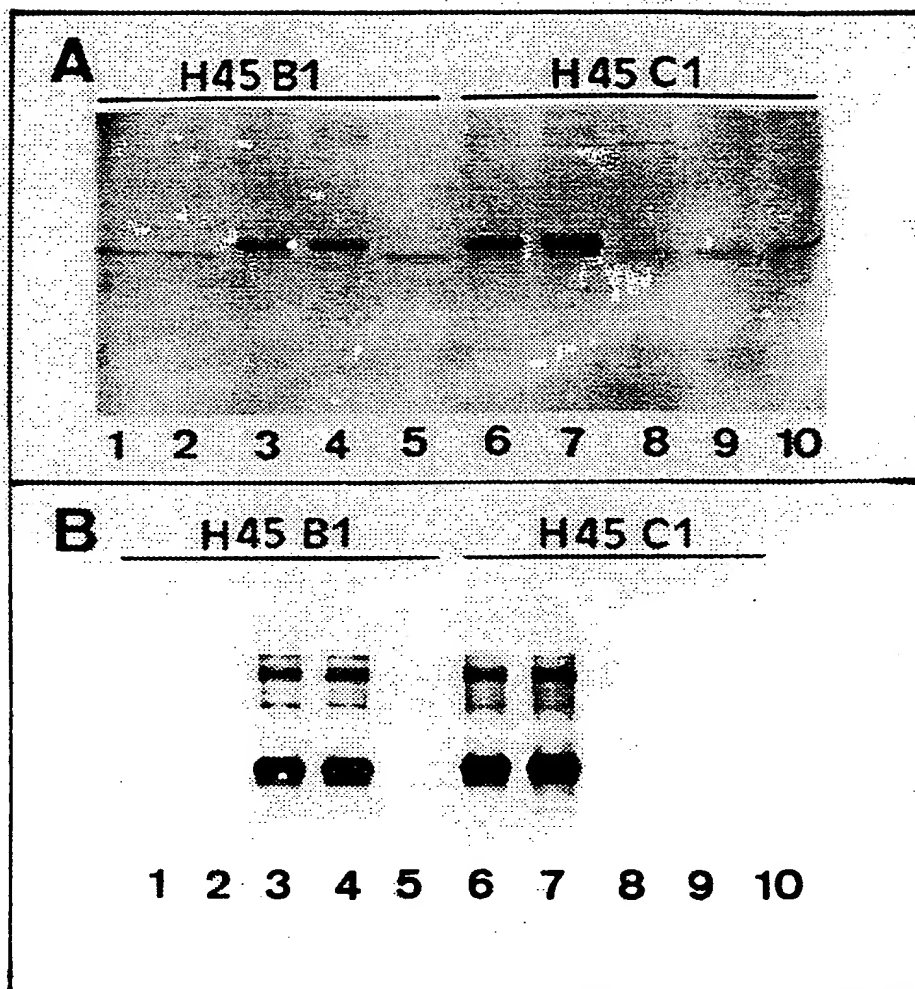


FIGURE 14



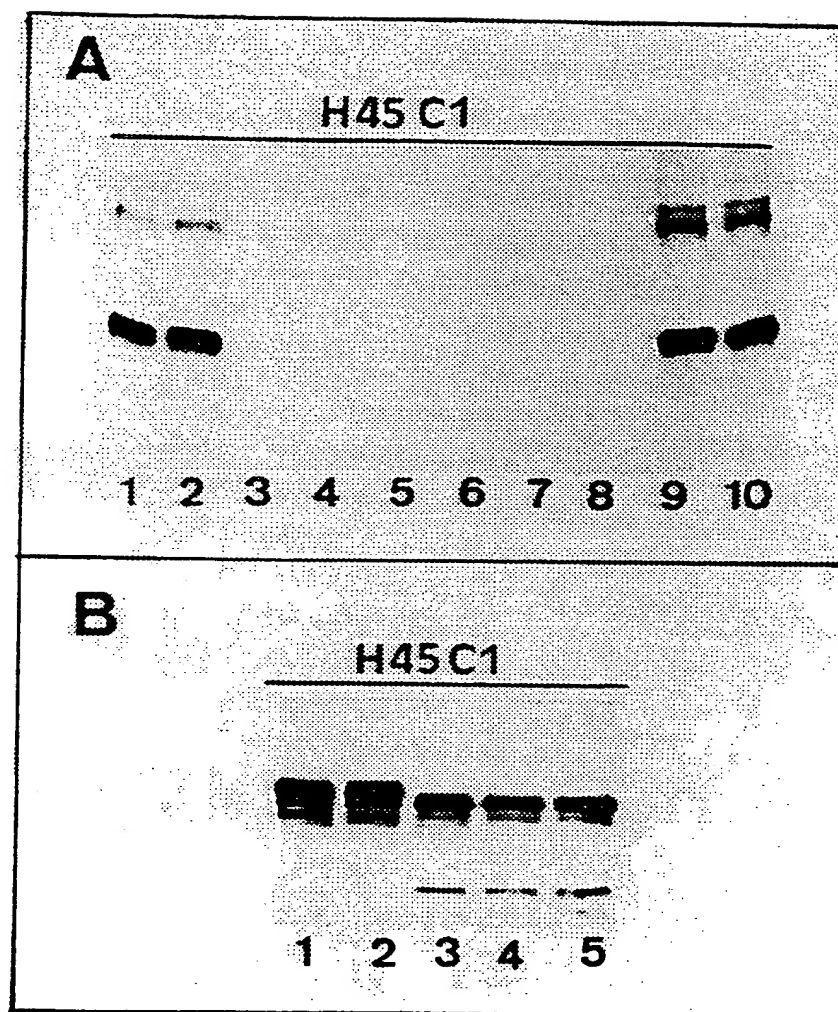


FIGURE 15

